ON MIND AND CONSCIOUSNESS

Selected papers from the MiCon2002 Conference

Edited by

Chhanda Chakraborti
Manas K. Mandal
Rimi B. Chatterjee

INDIAN INSTITUTE OF ADVANCED STUDY, SHIMLA
AND
DEPARTMENT OF HUMANITIES & SOCIAL SCIENCES
IIT KHARAGPUR

2003

First published 2003

Published by
The Secretary
Indian Institute of Advanced Study,
Rashtrapati Nivas, Shimla

Typeset at
Indian Institute of Advanced Study
Shimla

Printed at
Glorious Printers, Delhi

Acknowledgements

A successful conference is the result of the concerted efforts of many. We cannot mention all those we want to thank individually. However, we must mention the following organizations for their generous and timely support for our conference:

- Indian Institute of Advanced study, Shimla, India
- Infinity Foundation, USA
- Indian Council of Philosophical Research, India
- Council for Scientific and Industrial Research, India
- Indian Council of Medical Research, India.
- Indian Institute of Technology, Kharagpur, India

In addition to their support, the Infinity Foundation sponsored an invited session comprising American and Indian scholars from the fields of philosophy and cognitive science. We express our thanks to Infinity Foundation for making it possible for Dr Stephen Phillips, Dr Sangeetha Menon, Sanjay Chandrasekharan and Professor A.H. Khan to participate at our conference.

On behalf of the MiCon2002 Organizing Committee, we would like to express our special thanks to the Director of the Indian Institute of Advanced Study, Shimla, for agreeing to publish this book in addition to their generous support for the conference.

Foreword

The Indian Institute of Advanced Study, Shimla has organised an International Conference on Mind and Consciousness in collaboration with Indian Institute of Technology, Kharagpur on 9-11 January, 2002. Various aspects of the theme were discussed by scientists, philosophers, historians and others in this conference. Many problems and questions in respect of consciousness were deliberated in this conference. The question of dialogue between science and spirituality also emerged in this scholarly gathering. The present volume consists of selected papers presented at MiCon 2002.

I felicitate the editors of this volume for the active cooperation and the hard labour in producing the proceedings of this conference at an early date. I have no doubt that the present volume will be welcomed by the scholars interested in the study of mind and consciousness.

New Delhi

V.C. Srivastava
Director

Contents

PERSPECTIVES FROM WESTERN PHILOSOPHY

COMBINATORIAL PERSPECTIVES

EXPLORING MIND AND CONSCIOUSNESS: EMPIRICAL APPROACHES

LANGUAGE AND THOUGHT

Introduction

This book grew out of the international conference entitled Mind and Consciousness: Various Approaches, held between January 9-11, 2002 to exchange views, findings, and ideas on the topics of mind and consciousness across the disciplines and borders. The conference was organized by the Department of Humanities and Social Sciences, Indian Institute of Technology (IIT), Kharagpur, India, at the Institute as part of its Golden Jubilee celebrations.

As may be apparent from the representative papers in this book, this conference, which was the first of its kind at IIT Kharagpur, attracted notable scholars and researchers from both India and abroad. Delegates from abroad represented academic organizations from countries such as Australia, Belgium, Canada, Finland, Japan, Mexico, Sweden, Turkey, Hong Kong, UK, and USA. Participants from India, including research scholars, represented thirteen universities and seven institutes. We wanted to bring together researchers from various disciplines converging on the study of mind and consciousness and blending traditional cultural paradigms. We had amongst us scholars belonging to different disciplines; namely, philosophy, psychology, neuroscience, cognitive science, psychiatry, linguistics, physics and computer science.

The objectives of this conference were: to promote multi-disciplinary research on mind and consciousness, to work for a better and more holistic understanding of these concepts, and to explore whether Indian theoretical traditions can offer useful insights and solutions. We hope that the selected papers in this volume will help somewhat towards fulfilling these objectives.

With such a vast quantity of varied material, it was not possible to envisage a comprehensive volume of proceedings. We have therefore selected those papers which we thought have been particularly successful in making innovative or insightful suggestions in a representative area along the lines of the aforementioned objectives. Since Indian theoretical traditions are rich in expositions on mind and consciousness, a special section in this book is devoted to the papers which discuss Indian philosophical insights.

The book opens with three papers delivered as plenary speeches. The plenary speakers were chosen to set the tone of the conference and to outline the possibility of various multidisciplinary research directions. The next section, the largest in this book, is a medley of different theoretical deliberations converging on mind and consciousness. This section on the theoretical approaches is complemented by another, comprising papers with a distinctive experimental approach. Finally, a separate section has been created for papers on the interface between language and thought.

Among the theoretical papers, first subsection emphasizes the the Indian philosophical approach. Others, treating important and contemporary issues in the Western philosophy of mind, form the next subsection. The third subsection contains those papers which have tried to deal with contemporary issues from one philosophical tradition or discipline from the point of view of another.

THE PAPERS

Professor E. C. G. Sudarshan is a Nobel-nominated physicist of wide repute with a keen interest in philosophy. Through his experienced eyes, he acquaints us with what he visualizes as the possibility of a hard-core 'science of consciousness'; complete with suggested scope for experimentation. On the other hand, Professor Kapila Vatsyayana, doyen of Indian art and culture, speaks in her inimitable style about systems of thought in India which have employed research paradigms other than the typical empirical kinds. The latent argument is that there are other viable conceptualization techniques, more holistic than separatist in their spirit, for deliberation upon topics such as mind and consciousness. Professor P N Tandon is a well-known brain-surgeon. His paper exhibits his long-standing absorption in questions of mind and consciousness and shares with us the research possibilities which have opened up only recently due to the advances in the field of neuroscience.

Stephen Phillips, a well-known scholar in Indian philosophy, has considered the complex mind-body problem from the perspectives of three very different Indian philosophical standpoints: Sankara's Advaita Vedanta, Gangesa's Navya Nyaya and Aurobindo's Theistic Monism. His paper examines the way in which each of these schools of thought has tried to deal with the mind-body problem. The theory of self-illuminating consciousness that knows itself 'non-dualistically' is a central concept of Sankara's Advaita Vedanta. Philips claims

that by denying the possibility of access to consciousness externally or from a third-person point of view, Advaita has tried to avoid the mind-body problem. However, when it comes to .the issue of authentication of the pronouncement of consciousness on itself, Advaita settles the issue by making the self-illuminating consciousness self-authenticating also. This theory of self-authentication is what Gangesa's Navya Nyaya system finds untenable. The pluralistic Navya-Nyaya system, however, rejects the self-authentication theory and suggests that mind is not distinct from body. In the process, it claims that thus the problem of interrelation between mind and body in terms of causality can be avoided. Phillips argues that even Navya-Nyaya has its own share of mind-body problems in that of 'sensory connectivity'. According to Phillips, Sri Aurobindo, in his interpretation of this 'sensory connectivity' avoids the problem that the Nyaya school faces and has been able to defend a theory of consciousness that is intrinsically self-illuminating, yet is distinct from the Advaita position.

Sangeetha Menon's article proposes a fresh look at the subjective nature of consciousness from Indian epistemology and especially from Indian dramaturgy. She contends that clues found in these alternative perspectives could enrich our understanding, particularly of the unification of fragmentary experiences in terms of an 'I'. 'Consciousness considered as moral and the first person' by Indrani Sanyal looks at the way in which moral choices are made by individuals. As society is composed of a plurality of persons, each with his own aims, interests and conception of the good, various principles have become competing paradigms of 'ought' statements. With the advancement in technology and science, newer 'ought' and 'ought-not' questions are emerging. Sanyal examines the link between these ought questions and consciousness, and propounds an argument based on the analysis of voluntary activity from Nyaya-Vaisesika thought, which is carried out from the first person perspective of the doer of the action.

The section on Western philosophy of mind begins with William Johnston's contribution. In his paper, Johnston has taken the famous 'hard problem' (the claim that consciousness is what makes the mind-body problem really intractable) as the central theme and has looked at both sides of the problem. In his classic article 'What it is like to be a bat', Nagel identified 'consciousness' as the real reason behind the intractability of the mind-body problem. Johnston takes up on

one hand Nagel's view along with the positions of other philosophers, such as Jackson, Kripke, McGinn, Searle, and Chalmers, who have contributed to the gradual development of the thesis of the 'hard problem' from Nagel's comment, and the position of Patricia Smith Churchland on the other. Johnston sympathizes more with Churchland's argument that our inability to imagine a phenomenon with our limited existing knowledge-base has very little to offer as a premise to reach any kind of cogent conclusion about the objective nature of the phenomenon itself. The 'intractability' of consciousness, thus, in Johnston's view, is rooted more in our perception of the problem than in reality. However, he concedes that accepting the limitation of our knowledge-base does not ultimately make the problem of consciousness any more tractable for us.

The next three articles are reflections on contemporary issues in Western philosophy of mind. Steffen Borge has raised an interesting question about John Searle's verdict on the 'Chinese Room'. Borge argues that if we shift our stance from a metaphysical one, as Searle has preferred, to an epistemological standpoint and ask whether we have any reasonable epistemic grounds for ascribing 'mind' to a machine (as in the Chinese Room), then the conclusion that Searle drew from the Chinese Room does not seem so inevitable. Closely connected with the 'other minds' problem, Borge's proposal raises hopes for research on AI. Valtteri Arstila assesses the recent efforts of Graham and Horgan to bolster the Knowledge Argument originally developed by Frank Jackson. The Knowledge Argument tries to show the inadequacy of physicalism as the only metaphysical framework. It is based on a thought experiment in which Mary, a neuroscientist who knows all the physical facts that there are to know about vision but has never had any color experiences herself, is claimed to have a new, additional fact in her experience when she first experiences a color, say, red. There have been many versions, and restatements of this argument and just as many critical assaults on it. Arstila considers a recent redefinition of the argument and argues that the defense fails, as it is based on weak theoretical grounds. Sven Walter proposes a way to save the independent status of psychology and its specialized study of mind and the mental against the onslaught of arguments such as 'Multiple realizalibity of mental states' (by Putnam, Fodor), which collectively threaten to reduce mental properties to bare physical features.

Dominic Murphy, on the other hand, has posed an important and timely question about the direction in which psychiatry seems to be going. He points out that psychiatry recently has become more an extension of neuroscience, as the dominant trend is to treat mental illnesses as neuronal dysfunctions, and ultimately as at par with any other physical illnesses that are medically treatable. At this juncture, where distinctions hitherto considered important are being blurred, Murphy feels that psychiatry as a discipline needs to reevaluate its methodology, as well as its operational assumptions and conceptual foundation. as therapy may not be able to access or manipulate the factors identified as causes by neuroscience. In 'A definition of consciousness and some aspects of it as a physics problem', Tulsi Dass attempts to give a definition of consciousness that incorporates common notions of it, allows for gradations of consciousness in various lifeforms and can serve as a concrete goal for scientific explanation. The work builds on a research project by Dass and Manoj Srivastava, which proposes a thermodynamics-like formalism for the mental states. James F. Perry proposes that human thought and action may be stratified into three levels of responsibility: random, routine, and reflective. Perry argues that the reflective level has the distinction of becoming the highest form of human action and should be encouraged during the early level of schooling.

In the context of the mind-body problem, the concepts of supervenience and multiple realizability are important explanatory tools. Sanjay Chandrasekharan, in his 'Varieties of supervenience', traces two common versions of the supervenience relation to argue that one of them is ineffectual while the other, though effective, clashes with the notion of 'multiple realizability', which is pivotal for research on AI. He contends that a fresh perspective, such as the intervention of the concept of '*swasamvedana*' from Indian philosophy, can resolve this conflict. Saunders and Brakel bring up the issue that there is no consensus on the defining characteristics of 'consciousness'. Considering the fact that 'consciousness' is not a natural kind, they contend that perhaps a pluralistic understanding of 'consciousnesses' would be more appropriate. They cite historical and anthropological studies to argue that although there are 'family resemblances' among various notions, there is no one 'given' common idea of consciousness. Thus, they question the veracity of the assumption that neuroscientific explanation will one day reveal

the truth about consciousness, as according to them there is no unique referent to the term 'consciousness'.

Kundan deals with the problem associated with objectivity by drawing findings of quantum physics and from Kuhn's analysis of the history of science. Such findings suggest that there exists a world separate from the observer and that the nature of reality is relative with respect to individuals, time and paradigms. A resolution of the paradox of absolute and relative truth is attempted in this paper by using the arguments of the Madhyamika school of Mahayana Buddhism, the insights of Sri Aurobindo and the principles of yogic methodology. 'Person and boundedness in Wittgenstein and Tagore: positioning artificial intelligence' by A.H. Khan examines the concept of the person in strong psychological artificial intelligence and argues that central to the idea of person is the notion of boundedness. Khan revisits the history of the concept to understand the phenomenon of personhood, then turns to Wittgenstein and Tagore as resources to explain boundedness and person.

Among the empirical papers, G.C. Ray looks at consciousness in exalted or epileptic states through his long study of EEG results obtained from subjects, with special reference to transcendent experiences. Changes in the EEG readings in the form of isolated spikes or trains of spikes are observed during the 'clouding' of consciousness in certain medical conditions as well as meditation. Ray looks at the level of complexity of the human nervous system functioning under transcendence and suggests that in this state some of the subsystems in the human body may be 'switched off' or non-functional. He also deals with heart rate variability (HRV) and analyses the sympathetic burst during these states in the autonomic nervous system from the characteristics of HRV spectra. Ray then puts forward evidence from the memoirs of Ramakrishna in support of his contention that the neurological state of transcendence produces in the untrained human body the physical manifestations of a panic reaction, and that experienced meditators 'teach' their bodies to endure and explore this reaction. Brinck and Gardenfors report their observation of comparative co-operative behavior among humans and apes, two closely related species on the evolutionary scale. The paper indicates that through the study of the important differences that the humans exhibit, a better understanding of human conscious behavior can emerge.

Shobini Rao tries to explain consciousness from the perspective

of neuropsychology. She claims that neuropsychology may act as the bridge between the two domains of discourse; the phenomenal account of consciousness from Philosophy and the third person account of consciousness from the physical sciences. For instance, she cites clinical studies of brain-damaged patients and experimental studies of functional brain imaging which, she claims, have found that qualia arise from reconstruction of the sensory world by the brain. Clinical studies such as these, she argues, help to eradicate the 'mystery' surrounding consciousness, while augmenting the arguments from Philosophy.

B.N. Patnaik shows why generative linguistics is yet to come up with meaningful theories on the issue of the connection between language and thought. It has always been believed that language and thought are integrally related: language is the vehicle of thought. Generative linguistics does not deny this, yet contrary to expectations it has not been able so far to offer any interesting models. Patnaik finds that the problem is the lack of clarity over what the internal structure of a thought is and the relation between concepts and words; as a result there is no good theory of how a thought is mapped on to a lexical string, nor of how individual concepts are organized in the repertoire of concepts.

'Recognition' by Supriya Chaudhuri explores the place of cognition and recognition in plot formation in literature, drawing on formulations by theorists including Freud, Deleuze, Aristotle and Plato. Human knowledge is, for Plato, a form of remembering, while Aristotle speaks of that moment of recognition in tragedy when the protagonist comes to realize what is already known or indeed assumed in the tragic situation, so that the protagonist appears to proceed from ignorance to knowledge, yet his ignorance is founded upon a blinding of the self to facts already known. Chaudhuri argues that the literary text must be founded on this structure of loss at one level for recovery to occur on another. She examines some of the implications of this process of recovery and re-enactment and asks what the act of knowing involves. Finally, she brings these considerations to bear upon a theory of the literary text.

Paula Arizpe offers an analysis of Jaspers's existential methodology in connection with truth as communicative praxis. In this methodology, transcendence is an existential striving to experience or encompass the world in its totality.

Rimi B. Chatterjee looks at the problem of the representation of

reality in art. Epistemology privileges truth in an objective sense while the truth of art is, or seems to be, subjective. This division appears to set an insurmountable barrier between the pursuit of truth as it is found in the 'hard' sciences and that which the artist engages in. In modern society the right to tell the 'truth' as it appears in art is a problematic one, not least because conflicts arise between state power and the voices of individual artists. However, theorists of literature have largely ignored these questions, preferring to insulate literature's construction of the text from interpretations emanating from outside academia.

We hope that the volume will contribute to present debates on this subject and provide impetus to further research.

May, 2002
Kharagpur, India

CHHANDA CHAKRABORTI
MANAS K. MANDAL
RIMI B. CHATTERJEE

PLENARY LECTURES

Science and Consciousness

E.C.G. SUDARSHAN

The rapid progress in physical sciences has changed our concepts of matter, time, space, and energy. The advances in biology have shown the molecular basis for life forms. The human brain and nervous system have been thoroughly investigated leading to the beginnings of understanding of brain function. Many scientists (and their followers) believe that mind is but a function of the brain and consciousness is a loose anthropomorphic concept.

There are reasons to critically examine this view. In discovery, in insight, in devotion, in supreme aesthetic enjoyment, we find our awareness to be more or less unconnected with ordinary brain function. Where am I when I make a discovery?

The alternative point of view is to consider consciousness to be primary and brain function and sense functioning to be secondary. Such a view cannot be established by disputation, but only by the sure method of experimental science: one should relate one's experience to this enquiry. If a person has no feeling for music or painting, he (or she) could argue that such things do not exist, or if they exist, they are simply superfluous action.

Witnessing awareness (*sakshi bhava*) is the one a scientist is most aware of: and the awareness in this context is extraordinary but experienced.

These various views will be compared and contrasted.

Introduction

The threshold of the twenty-first century finds us with a scientific basis for understanding material phenomena whether it is for living or nonliving systems, whether it is for the very small or the intermediate or very large structures. For the study of matter we have a fairly quantitative understanding in terms of chemistry, atomic physics, and quantum mechanics. When larger structures are involved we need the principles of engineering, geophysics, astrophysics and cosmology. For living systems the past century has

contributed a molecular basis for their structure and functioning, including the genetic basis for invariance and the possibility of mutations. Most biologists are comfortable with the statement that molecular biology now provides a quantitative framework for their science. As to the nature of matter in its basic nature quantum mechanics, quantum field theory and possibly new developments like M-theory provide the basis, though we have been forced to accept dramatic revisions of our concepts of structure, dynamics and even the very nature of space, time and causality. Unexpected relations between the very small and the very large have been revealed and today cosmology and particle physics including relativity and quantum theory are closely related.

Condensed matter exhibits many new features necessitating new methods of study but the basis of all these studies is quantum mechanics. The advances in this field are such that not only can one understand the behavior of most materials, but we can also make "designer materials" with properties that make them useful in computation and transformation processing. One of the earliest of such materials, the Bose condensate, followed from the principles of the new statistics that Bose and Einstein formulated even before quantum mechanics itself. The behavior of Bose condensates is currently occupying a number of gifted experimentalists and theorists.

New Frontier for Science

Are there new frontiers? Yes, of course, the nonlinear dynamics of fluids and solids is a young discipline, but one which deals with very familiar processes like boiling, crystallization, meandering of rivers, formation of snowflakes and the draining of a bathtub, also, in general chaos and pattern formation. Despite the complexity of the processes and products, we believe the laws of matter in the small, namely quantum mechanics and particle physics form the basis. But as the complexity of the system increases quantitatively new properties may emerge, so matter in the bulk would need the study of complex systems to supplement the basic microscopic laws.

One of the very complex structures that we need to study is the human brain. Its neurophysiology and electrochemistry are among the most complex structures that we encounter. Painstaking research has elucidated much, both to the mechanism and the functioning of the brain. We have some understanding of neural pathways, and neural networks and many stages of the processing of "signals." We could map domains in the brain with memory storage, language and sensations though every theory for brain functioning seems to be incomplete. The brain is the organ most people associate with the mind. Some people even tend to identify mind to be the function of the brain.

The mind-brain interface

This interconnection is both an opportunity and a complication in the study of the brain and its functioning. We think we have volition and initiate action by the body via the brain. Unless we want to take the extreme position that volition is illusory and the brain tells the mind to act as if it has volition; we have here a complicated material object being activated by something that is non-material which would imply that the physical world is not closed. This, if it were so, is a startling state of affairs.

Undoubtedly the brain tissue acts on the mind. Not only do sensations get communicated to the brain, but also suitable chemicals in minute quantities in the synapses can alter the mind. Does this mean that the mind is not autonomous, but is subject to action by physical agents?

On the positive side, this interplay allows us to study aspects of the mind in terms of the brain and aspects of the brain in terms of the mind.

The study of mind itself is a vast subject, which is partially dealt with by standard psychology. Mechanisms for the various modalities like volition, recall, purposive action, deductive and inductive thoughts are in the province of physiological psychology. However the agency called the mind seems not to be able to discriminate between these and other domains of inner experience. Contemporary science accepts and even demands controlled experimentation as the modality of knowledge of the physical universe. But why not explore the experimental and experiential processes to get more reliable knowledge on which to make a science of the inner self?

In the experimental methods of physical science great store is put in the fact that it is public knowledge on which the science is built. The individual's variability of interfacing observations is

thought to be eliminated by using properly calibrated meters with their digital or printer readings. Should we not attempt to do the same for mental processes? Psychometry does precisely this, but then psychometry can access only a very small part of the inner domain of experience.

Levels of the mind

Much confusion and miscommunication arises because "mind" is taken as a "one-size fits all" agency in many of these discussions. This is one step better than the position that mind and the brain are inseparable. The hallmark of mind in this context is thought to be which "I think, therefore I am" which sums up this reductionist view. The faculty of discrimination, the sense of individuation, and the creative intelligence are left out or clubbed along with the thought modality. In the Indian tradition we call them *buddhi, ahamkara, citta* and the level of ordinary thought as *manas*. We shall keep these distinctions in the sequel.

In any laboratory experiment in the physical sciences, one tries to produce a controlled and sheltered space and time, say a laboratory. In case of the study of inner experience one should do the same. We should seek isolating shelter, which is called *kaivalya* in the practice of the discipline of mindfulness. One has to take elaborate preparations and dedicated practice to remove the chance variability of the inner experience. The functioning at these various levels would then be seen as distinct, just as inorganic chemistry, atomic physics and nuclear physics are levels that we do not mix up if we are serious in the study of matter. They all pertain to matter but they are appropriate for restricted domains of physics.

Are there not other schemes by which our inner experience can be analyzed? One big problem is to overcome the possibility of volition. This spontaneous, non-causal activity cannot be encompassed in terms of standard psychology. As a solution Freudian psychology postulated levels of the mind which are not directly accessible to a person. By postulating powerful agencies and mechanisms to the subconscious mind an attempt is made to restore causality to the inner world. After all, was this not the way that physics took care of electromagnetism? Is this not a way to expand the domain of the science?

But we must recall that physics has no category for 'volition'; no free will for physical defects. This is so notwithstanding the uncertainty principle, (which is only an expression of the inappropriateness of the usual language). In any case uncertainty is not choice. So if we accept our experience of freedom in volition we must look for a new property not encountered in physics.

Quantum computing

There used to be a time when many people found the electrochemical and neural networks to be indications of the nature of the brain as a giant supercomputer. Despite the charm of their view one could easily have foreseen the inadequacies in such a view. All digital computers operate with arithmetic and Boolean algebra, either yes or no (bits) as the elementary bit of information. We recognize that the mind, if not the brain, is used to ambiguity, superpositions and nonarithmetical nonsequential processing.

During the closing decades of the past century people came to regards two-level quantum systems (qubits) as the building blocks of a computation system. The passage from bits to qubits enables us to make use of the many possibilities of coherent super-positions of the two basic states. These may be realized in terms of two-level atoms. Coupling of qubits using logic gates can lead to new methods of processing information which are sometimes superior to the usual bit-processing digital computers. All such quantum computations can be made reversible. Two useful quantum algorithms are the one discovered by Peter Shore, useful in factorizing huge numbers, and the one discovered by Len Grover for picking out a particular entry from a huge database.

So far in quantum computations, the principle of superposition of quantum states is utilized for bipartite and mere complex systems; there is another characteristic property, that of quantum entanglement. This property can make quantum information available at widely separated points without any propagating effect going from one point to the other. This has been used for 'quantum teleportation' for secure communication between two individuals without the signal being intercepted or copied by any intelligence service during the transmission.

We shall have occasion to refer to quantum entanglement and

teleportation in the context of communication of creative insights both in art and in science.

Consciousness

This new dimension is that of consciousness which encompasses all the function of the mind but contains new potentialities. We have already come across freedom in the form of volition and it is also evident in creativity whether it is of an artifact, a performance or a new perspective on known facts. In a world where both the external and internal realms are impermanent and changeable the only entity that is invariable is consciousness. And finally we much recognize the dimension of pure happiness at which time neither the internal world or the external world makes an impression. Experientially they cease to exist.

Is it possible to have a science in which consciousness is included? The natural attempt is to include this additional aspect as a new constituent of a physical-mental-consciousness complex. If we have such a science it cannot obey the kind of laws with which we are familiar. Another possibility is to consider consciousness as primary and the other mental levels as emanating from consciousness and subsiding in it.

Two questions arise in this case. The first is to ask whether the choice of consciousness (*sankalpa*) can determine the functioning of the mind: the element of freedom inherent in consciousness would then percolate to the grosser level functioning including the domain of thought. We should search for such manifestations in a scientific manner. The mind becomes the laboratory. The proper sheltering of the mind-laboratory is to be brought about by an appropriate protocol.

The other question is the old question whether the mind can affect matter, or in a revised form, can consciousness affect matter? If it can, can it be scientifically demonstrated by careful experiments with appropriate protocols? People have done experiments to test whether human intention can affect a random number generated, or the rate of decay of an unstable physical system. More recently William Tiller and his collaborators have shown positive results in a number of experiments. They declass that not only are their results in a number of physical phenomena reproducible, but that they can condition standard physical laboratory equipment and ship it to other laboratories for study by other scientists. If this gets generally

accepted one could say that the choice characteristic of consciousness spills over into matter.

Is consciousness singular or plural?

This in turn raises another question. How different are "your" consciousness and "my" consciousness? Can the free choice available to your consciousness limit the freedom of my consciousness? One way to avoid the predicament is the traditional wisdom that consciousness is *One*. When the individual limits it we term it "your" consciousness or "mine." It would, in this view, be more precise to say consciousness in you or me.

Distinct but related views of consciousness have been postulated by C. Jung. He recognized that the universality of symbols in art and in dreams across societies and epochs may have a natural explanation in terms of a collective unconscious from which each of us draw. This view is opposed to Sigmund Freud's model in which each person has his or her own sub-conscious mind which contains the symbols and signs for dreams. Later writers have elaborated this view to include art and literature.

Can consciousness condition a laboratory?

When we consider the mind-matter complex one has to take into account extraordinary transmissions of information that have been reported. In these cases the information does not seem to be propagated but just is. In normal propagation the information signal becomes weaker as instances go, but in these extraordinary processes it does not seem to be so. Shielding and opaque screens do not affect there. It is as if the information is there just to be picked up, rather than get these from one point in space and time to another. In the brief account of quantum computing we came across such shared information in quantum systems. Rather than discard all such reports of distant reception of information as unreliable we should show them as opportunities for careful experimentation. We must recall that at the beginning of the twentieth century there were a number of small effects, which could not be integrated with the theoretical framework. But eventually it was the framework that changed, from "classical" to "quantum", rather than the 'incomprehensible' physical effects.

Creativity and aesthetics

One context in which consciousness acts on the mind and, in turn, on the public domain, is that of creativity. Just like "choice", creativity also does not find a place in contemporary science, despite the fact that many scientists have direct experience of creativity. Creativity in science or in art eventually manifests in the public domain. The creative scientist (or artist) is the best authority on how the process manifests. There is the final expression stage (*vykhari*), the processing stage (*madhyama*), and the stage at which the idea is perceived (*paśyanti*); but each creative person is aware that there is an unfathomable level (*para*) from which the creative insight emerges. It is not common for scientists to trace the stages of creativity, but more such accounts would undoubtedly clarify the process for us.

Not only are these stages part of the emergence of creativity but a reversed sequence is there for artistic enjoyment. The performance (or the artifact) is in the public domain, but not everyone gets the same enjoyment. Once the enjoyment is actualized, there is a private perception of the creator's idea, but that perception is definitely private. The culmination is in happiness, calm undisturbed consciousness. The artistic element going through the first three states is often referred to as *rasa* according to Barlingay. The rasa has the texture of a fluid and the flow depends upon the audience. (In scientific creativity we do not have any study of aesthetics; and we ought to have it.) But individual scientists can vouch for these stages of passage from the uncharted to the charted domains and both the creator and the consumer see these stages as pure happiness and peace.

One question that may arise is the following: can a creative idea be transmitted from one person to another without the public domain activities? Can a master scientist (or artist) impart knowledge without the use of the final stage? My years as a student and then as a master in science convince me that it is done. It is also in accordance with most traditions. But for a scientific study we need more data and more careful observations.

Science is experience driven often by deliberate experimentation but also by passive observation, but theory is the trailblazer: even an experiment that disagrees with existing theory points to an improved theory which makes the experimental findings become a coherent

body. This method is not inapplicable to the domain of consciousness; it is only not explicitly recognized. A scientist or a saint is a disinterested but diligent observer of his or her own experience. Rather than be in antagonism and be adversaries, those who study the natural sciences and those who study consciousness can be cooperative and open.

Concluding remarks

The twentieth century found itself with many new frontiers: the constitution of matter, the architecture of life, the problems of a physical cosmology and texture of complex systems including the brain. The twenty-first century finds consciousness and its relation to the brain to be the new frontier. Our experience with quantum mechanics, brain physiology and quantum computing should help in this new domain, but we are bound to have startling surprises.

Mind and Consciousness
Opportunities for Scientific Studies

P. N. TANDON

You may feel a little let down by what I want to talk about at this prestigious conference. I, by training and by profession, was a neuro surgeon till a few years ago who thought I knew something about the brain but also realized that there is so much more to learn about the brain than what we knew. After handling the brain for 30-40 years to deal with its physical problems caused by injury or disease progressively over the years, I got interested in the complex problem of brain-mind relations in brain and consciousness and I still continue to be a student of this fascinating subject. Therefore, one of the purposes of coming here was to learn from such distinguished people. However, since I have been asked to do so, I take the opportunity to share with you some of my thoughts on the subject. I will try to say where we started and where we are in our efforts to understand this complex subject.

It is not surprising that mind has been a subject of study for eternity. As far as we have recorded literature on any civilization, there have been references to the study of mind. Many people believe, and certainly most of the Western scientists start their talk about this complex subject by quoting the Greek scholars Galen, Plato, Aristotle, Hippocrates and others as if they were the first to study this subject. They do not realize that this has been a subject of great intuitive and logical studies in this country, India. We lost track of much of this valuable knowledge but in this audience, there are scholars who have been studying these resources and bringing out the nuances of these for people like us. One of them is Professor Mukhopadhyay for whom form I have great admiration; then we have Dr. Sangeetha Menon. Just to mention a few examples for the Indian resources, the *Taittiriyopanishad* classified consciousness at different planes: the physical consciousness, the psychic consciousness and the beatitudinous consciousness. Professor Srivastava,

Director, IIAS Shimla, and Professor E.C.G. Sudarshan, in their talks in the inaugural session of this conference, have already referred to some of this. In contrast to that, at a very basic level, which I think the modern scientists will have no difficulty in accepting, is the *Mandukopanishad* which describes four states of consciousness –*jagriti*, that is awake, *swapna*, that is dreaming, *Susupti*, that is deep sleep and *Turiya* or the non-dual thoughtless state. Today, the modern scientific community finds it difficult to comprehend the concept of the state of *Turiya*. Whether we agree with these concepts or not is a different issue, but the fact that they thought of all this goes to reflect how much intellectual exercise had gone on in this area. Again the literature is so vast and it is far beyond me even to attempt to briefly summarize the essence of it. I only wish to point out that it has been a subject of study for more than several millennia.

In the context of mind, one of the most quoted authors of the West is Descartes, who considered the body and the soul as two distinct entities, and went so far as to localize the point of interaction between the two to the pineal gland. I do not know of anything that has been more highly quoted than this single diagram of his interpretation (see picture. 1) of the brain and mind relationship in which the mind was supposed to be in the pineal body and the brain only concerned with sensations. Notwithstanding the utter improbability of such a hypothesis, it introduced the concept of dualism and a trend for a reductionist approach which has dominated so much the thinking of Western philosophers and scientists alike.

In any case the most important question that arises in one's mind is: if consciousness is the function of the brain, does it have a seat in the brain? Can brain mechanisms explain consciousness in its entirety? Now, this word " entirety", is what caused the greatest trouble for people like us. What is entirety? Is entirety in the terms of what the philosopher thinks, what the religious sages think or only in terms of what little literate clinicians like me think? If we could just remove that word "entirety", I would easily say, yes, the brain mechanisms explain consciousness as I understand it. I will come to that in a second. There are other unresolved questions too, such as: Are consciousness and mind synonymous? What is mind? Is the brain the organ of mind? etc. Now if you take up an unsolved question and try to solve it, it is wise to convert it into simpler propositions. Therefore, having worked as a clinician and caring

for patients with brain disorders affecting their consciousness morning, evening and night, it became essential for us to develop a working definition of consciousness and its gradation. To us, the clinicians, consciousness is *the ability of the person to be aware of his/her environment, to be aware of himself or herself and to have volition.* These are three major components and this is how we tested our patients by checking whether he recognized the people around him, knew who he was, where he was, whether he remembered what day of the month it was, could recall his life-events from memory. Similarly if we asked him to lift his arm and lift his leg, was he able to do that. That was very simple and we quickly gave mathematical numbers to it so that we could predictably grade his state of consciousness.

It is not surprising that, in our attempts to resolve any problem at a given moment of human existence, we go only as far as our tools will permit. That is no doubt a very simplistic view. In any case initially we had no tools and therefore we used our brains and we had what is written in the *Upanishads* and *Vedas* and what Aristotle and Hippocrates talked about. These thinkers had no tools. They of course had their minds and their observations and their intuition and their ability to logically develop some hypotheses. Progressively, various sciences developed and we started to utilize them and study the same subject – mind and consciousness from different stand-points. Therefore, a whole group of people developed theories of mind and consciousness. There were those who believe that consciousness is a unitary state, and those who believed that the brain and mind and consciousness are two different things, and there were mystics who believed in only one consciousness that was the state of *Brahman* or *Ananda.* Surprisingly, all mystics, irrespective of the culture they belonged to, have described the same experience in different languages but with identical characteristics. That means there must be some commonality in this experience. Hence whether I believe in it or not, I would not like to dismiss this as irrelevant for any study on consciousness. However, as we moved from religion to philosophy, from philosophy to psychology, from psychology to neurology, from neurology to cognitive science and today to a combination of all these approaches to study, we find the emergence of behaviorist, functionalist, mentalist, pan-psychiatrist, parallelist and emergent materialists from time to time who provided their own interpretation of the subject. It only depicts what Prof. Sudarshan just mentioned; namely, that you can go only up to the limits of your capability at a given moment of time and that applies not just to the individuals but equally to collective intellectual bodies.

Before I discuss what constitutes consciousness, let me tell you what is mind to us. Mind is *what reflects the mental activity.* What is mental activity? Anything that goes on between the sensory input and motor output. I see somebody: that is a sensory input. But that I recognize him, that I have an emotional response associated with it, that I have will power to decide whether I shall go and shake hands with him or whether I will turn around and look the other way around, I then act accordingly; all that goes on is mental activity reflected in the mind. Campbell in 1980 concluded consciousness is not identical with brain or mind. It is that which witnesses changes which occur in the brain and mind. Surprisingly, a similar conclusion was arrived at after prolonged discussions between a most out-standing neurophysiologist of the last century, John Eccles, and a philosopher K.R. Popper. They sat together, debated for days, and brought out a book *The Self and Its Brain* (see Popper and Eccles 1977) and came to an astounding conclusion: that "the unity of conscious experience is provided by the self-consciousness mind and not by the neural machinery of the liaison areas of cerebral hemisphere". William James (1892), the father of the American psychology, said that consciousness is not a thing but a process. I think that is a correct statement even today. It is not a thing, it is not a material thing. He added that everybody knows what consciousness is until he tries to define it. And I can tell you that having devoted last 10 years specifically to the study of this subject, I have no better definition than what I gave you. But that is too simplistic, a definition as Prof. Mukhopadhyay would say does not cover its entirety. It does not cover its entirety because we are incapable of handling that entirety. Somebody who is more capable may like to join hands with us and see whether we can expand our horizons.

Let me attempt to elaborate a little further as to the various components of consciousness. Awareness, attention, cognition, discrimination, memory, responsiveness, volition are at least some of the definable constituents of the unitary state of consciousness. Consciousness is not attention or memory or any one of these individually including awareness. It is a combination of all of these and this is why it is so complex. Since they could not be quantified, earlier we had problems to study them scientifically. We have come to a stage where each individual component referred to can now be

studied quantitatively as I will illustrate during my talk.

Since this audience may have people who have no idea of what brain looks like, let me just show the picture of a human brain as exposed during an operation. The cerebral cortex is divided into frontal lobe, parietal lobe, occipital lobe, and temporal lobe. You can see this is the cerebellum, that is the brain stem. When you split the brain in two halves, this is the picture you get. These are the bundles of fibres which connect one half of the brain to the other half. I do not need to go into the details of all this. The building blocks of the brain are the nerve cells (neurons), supporting cells (glia) and fibre bundles. The neurons are interconnected to form neuronal circuits or networks. Each neuron is connected to 100 to 1000 other neurons. The junction between the neurons is not in continuity. There is always a gap between the projection from one neuron to the other neuron. This is called a synapse. Charles Sherrington, a Nobel Laureate neurophysiologist, said that *consciousness is a phenomenon of synaptic activity.* Many others believe today that consciousness is a phenomenon which is the result of the connecting networks functioning synchronously at a given time.

As late as 1966, and even later than that, Sperry (1966), a distinguished neurobiologist, eloquently stated that we were having problems with understanding consciousness, because the inner sensations, feelings, concepts, mental images and the like cannot be weighed or measured, photographed or spectrographed or chromatographed or otherwise recorded or dealt with objectively by any known scientific methodology. That is a little less than 40 years ago. I will try to illustrate that the situation has changed dramatically so that at least some of them can now be photographed, even spectrographed, though they may not be chromatographed.

Let me digress a bit. One of my teachers is Dr. Wilder Penfield from Montreal, who studied brains of conscious human beings during surgery for epilepsy and meticulously recorded every single response following gentle electrical stimulation. Towards the end of his scientific life, he thought that he would like to answer the question which has been a matter of great concern to everyone:– what is mind, are brain and mind the same or two different things? The last scientific book that he ever wrote was *The Mystery of Mind* (Penfield 1975). I will suggest that those who are interested in the subject should read it. Here I quote his conclusion:

> From my own part after years of striving to explain the mind on the basis of brain action alone, I have come to the conclusion that it is simpler (and far easier to be logical) that one adopts a hypothesis that our being does consist of two fundamental elements – the brain and the mind.

Wald (1979), another outstanding biologist, argued that consciousness is not a part of the universe of space and time, of observable and measurable quantity that is amenable to scientific investigation. You see this was as late as 1975 and 1980, still the feeling was that we cannot study it scientifically, and therefore we leave it to the philosophers who can understand it probably better than we do.

There were others who felt that it never will become possible to identify physically the presence or absence of consciousness, much less its contents. I can quote the names of many in this regard. One of our most outstanding, and most respected physicists, Dr. D.S. Kothari, said:

> atom or brain belongs to the external objective world of space-time and matter-energy. Self or mind belongs to the internal subjective world. The subjective (mental) world is not describable in terms of space-time and matter-energy concepts. Self or mind, therefore, lies outside natural science (Kothari 1980, pp.1-28).

This was in 1980. Twenty years ago, people were still thinking that they cannot tackle the subject scientifically. I, with all humility, tend to disagree with this and would like to illustrate in the next part of my presentation, where we are. As I mentioned at the outset, I am not, for one moment, saying that we have solved the problem of the mystery of mind or of the brain-mind relationship or the mind-consciousness relationship. But we now have the tools with which each of its components that I referred to earlier can at least be studied scientifically. On the basis of our own studies and the recently emerged new information, in my Presidential address to the Indian National Science Academy, I concluded that,

> On my own part I am convinced the mystery of consciousness is inexorably linked to the mystery of the brain, this mystery is more likely to be resolved, if ever it does get resolved, (that shows my uncertainty, my not being too confident) by a holistic approach to study the brain as it was becoming possible now through investigations of the human beings (Tandon 1993).

Three to four days ago, I gave a talk at the Indian Science Congress Association in a symposium entitled *From Molecules to Mind.* The title of my talk was 'Demystifying Mind'. I am trying to demystify, but neither trying to say that the mystery is solved nor trying to make light of the complexity of the problem. As a clinician, I saw so many clinical syndromes which have a bearing on the subject. I have just chosen a couple to illustrate. Patients with Korsakoff psychosis, for whom we know where the lesion is, a person has loss of memory. But if you meet one, you will find him a jolly good fellow, he talks to you, confabulates in answer to your questions. Five minutes later when you meet him, he asks you who you are and when you say that you met him just a few minutes ago, he exclaims that he has never met you in my life. This we call a cocktail-party syndrome. Even without a brain lesion under the influence of some toxic substances, one can behave as if he neither has awareness of self or his environment, but for all practical purposes he appears as a normal 'conscious' human being. As per the definition of consciousness that I referred to earlier, such a person cannot be considered to be conscious. Similarly a person with a lesion in the brainstem no bigger than half a centimeter in diameter has no perceptible voluntary movement in his body. When you talk to him, he listens, he sees you, but he has no means to responding to you and you take it for granted that he is unconscious. But the moment he recovers that capability, he can tell you everything that happened to him during that period when you thought he was not conscious. Thus his ability to control his expression voluntarily, which is one component of the working definition of consciousness, could be misinterpreted as loss of consciousness.

There are any number of such syndromes which create problems for establishing a fool-proof definition of consciousness. For example, global amnesia is another. During an episode of global amnesia a Professor giving a talk, like me here, suddenly starts fumbling, not knowing what he is doing and he even wonders aloud "what's the matter, what's the matter, what has happened". When questioned, what is wrong with him, he declares "I am alright". Informed that he was giving a lecture, he says, "No, I am not giving any lecture". Is he conscious or unconscious? At least he has no memory of all this even when he recovers. Study of patients with well-defined lesions in the brain has provided a wealth of information on various aspects of consciousness.

Two of the most classical cases are depicted here (slides shown). This person had a knife injury. The knife went through the face and the base of the skull into a very vital area of the brain, resulting in at least one aspect of his conscious functioning, specially in respect to recall of past memory being affected. This other patient who was operated upon for epilepsy had one temporal lobe including the hippocampus removed (as illustrated). I have removed hundreds of temporal lobes. We know removal of one temporal lobe does not result in any discernible deficit. Based on such assumptions, somebody removed both sides. This person had no capability of recording any on-going experience even though his IQ test was perfectly normal. His behavior appeared perfectly normal. He was a draftsman and he would go, sit and do his drafting work perfectly well but when you asked him after five minutes what he was doing there, he has no idea what he was doing. Was he conscious or unconscious? These are the kinds of questions that started to creep up in our mind. There are beautiful descriptions of these patients who have been studied endlessly by psychologists and imaging specialists using latest techniques of brain imaging. Analysing the data thus generated starts to provide new insights on brain function and its correlation with behaviour.

Yet another example. We had virtually removed one half of this young boy's brain (slides shown). He never lost his consciousness as a result of this. He was of course totally paralysed on the opposite side i.e. left side and he needed a support to walk. This shows, we could remove large chunks of brain without impairing consciousness. Thus one sided removal or injury to any area of the cerebral hemisphere, even whole of it, does not abolish conscious thinking. It may change the contents of awareness, interfere with voluntary act, render less effective planned action, deprive the patient of word symbols, but he still thinks and weeps and sometimes laughs also at his own self if not at others. Is he conscious or not? Another case of a little tumor in an area of brain and removing that tumor one had to go through the whole brain. There is a tiny little structure in the brain called the fornix and if by any mistake in removing this tumor one damages this structure the person loses all recent memory. It is now obvious that memory is not necessarily equivalent to consciousness but certainly is an important component of consciousness.

I am only trying to give these examples to illustrate how valuable

insights can be gained not only about brain function but about consciousness from study of such patients. In yet another way one could study the same phenomenon. During surgery for epilepsy on conscious patients under local anesthesia, gentle electrical stimulation of the temporal cortex in some patients activated the past record of the stream of consciousness. Limited to the duration of the electrical stimulation the patient re-lived all that he had been aware of in that earlier experience a record of which was laid down initially. No doubt this could not be achieved in every patient submitted to such stimulation. But the evidence provided by these few patients as if there was a cinematographic or video recording of past experiences in brain would strongly support the contention that the brain is an important organ responsible for our consciousness.

As a part of the management strategy, surgeons introduce electrodes into the human brain to record electrical activity (electro-corticography). These studies are carried out for treatment of a patient's disease but while doing that we utilize the opportunity to study the electrical activity, and response to stimulation. By these observations we can safeguard the interest of that patient and future patients, at the same time obtain valuable information about functional localization in specific regions of the brain. With more recent technology, similar information can be obtained from the surface without putting anything inside the brain, by doing electro-magnetographic recording. Thus recording can be obtained from well-defined areas of the brain even in the depth. Simultaneous recording can be done from many areas when the person is carrying out a voluntary act or a psychological exercise. This is done in real time, and during that period the patient's experience can be co-related to activity in specific regions of the brain.

During the last 15-20 years, CAT scan, magnetic resonance imaging (MRI scan), functional magnetic resonance studies (fMRI), Single Photon Emission Computerized Tomography (SPECT), magneto-encephalography (MEG), provide non-invasive methods of studying anatomy, physiology, chemistry of discrete regions of the brain in fully conscious, thinking and performing individuals. It is now becoming possible to do such studies even in infants.

Let me show some examples to illustrate how even higher mental function can now be studied. It has become possible to see which part of the brain is activated while the person is hearing words, or

when the person is seeing words, or when the person is speaking words and when the person is generating words. When a person is shown some nonsensical symbols of which nobody knows what they are, one part of the brain is active. Next when some English alphabets which cannot be pronounced as a word were shown another region was activated. When shown a word that has no meaning, the activation is at a different site. And yet another region is activated when a recognizable word is presented.

There are other ways of studying functional localization using fMRI. Utilising even more sophisticated modern technology of electro-magnetography, using SQIDs (Superconducting Quantum Interference Devices)on the surface of the skull, not on the brain, one can study in real time things that are happening in the brain during such processes as memorizing or doing mental mathematics.

In this manner it has been possible to delineate the parts of the brain that gets activated during anger, happiness, fear and sadness. Carried out under certain specific circumstances with inherent limitations, there are still technological problems, which need to be sorted out to improve these results. But there is a difference in localization between happiness and anger and for that matter between anger and fear. One can thus get information from a variety of techniques, not just one technique, for the same function because some give us very good anatomical resolution, whereas others give us functional information. Some can only be interpreted when we have collected data over several seconds, while others can give us data which are in real time. All this information needs to be integrated into one unified concept.

Even more complex behavioral traits have been studied like identifying the areas of the brain involved in persons addicted to cocaine. Consistently we find higher activity in the basifrontal region and medial aspect of the frontal region. The change after he has been treated is so obvious. Addiction is a very major societal problem today in the world. If we start to understand the basic mechanism underlying this, even if we leave aside the philosophical aspect of consciousness, and look at the practical aspect, we can derive immense benefits.

From the data that have been collected, we start to find out that the left hemisphere deals with reasoning, problem solving, spoken and written language, mathematics and science. The right hemisphere which is supposed to be more dominant in women

accounts for insight, imagination, art, awareness, hearing, melody and pattern recognition. Study of patients following surgical procedures that disconnect two halves of the cerebral hemispheres suggests that the two sides can be independently conscious.

The brain imaging studies of more complex mental functions in health and psychopathological states have revealed the areas of brain involved with emotions. The *following illustrations* demonstrate the areas of the brain showing activity in a patient who was in a panic state, or another one who was in a mamac stage of manic depressive psychosis quite different from when he was in a depressive phase. We are inclined, maybe wrongly, to believe that this part of the brain is linked to this particular aspect of a person's emotion or mental function.

Notwithstanding this voluminous information already available and being added to everyday, the question whether we are nearing identification of the neural mechanism underlying consciousness, leave aside mind, remains unanswered. We no doubt have piecemeal information about the various elements that we believe give rise to the unitary subjective experience of consciousness or self-awareness. Yet many questions remain unanswered. The foremost amongst these are,

i) Does demonstration of neural activity in a part of the brain reflect a mere association with a particular mental phenomenon or a cause for it?

ii) How is the activity in widely distributed areas of the brain even for simple sensory function integrated to provide a comprehensive experience?

iii) Are there specific mechanisms or specific circuits responsible for this integration?

iv) Can the same mechanisms explain phenomena like "mystic experiences", "hypnotism", "out-of-body experience" and such others so vividly described by large number of persons from diverse background?

No doubt there are many more such questions. The purpose of this presentation, I must acknowledge, was only to demonstrate the current scientific approaches used to study this subject and our present state of knowledge about them. I hope this presentation will stimulate scholars from other disciplines, gathered at this conference, to come together on this difficult journey to unravel

the mystery of mind and consciousness.

Take this very simple thing vision, which is the most studied perception system. You see in the eye, it goes to the brain, goes to the back of the brain. It does not go as a full total picture. Different parts of the brain are involved in different functions. For instance, colour for this part, for the depth of the subject, if the subject is moving the position that is the distance from you to the object and different part of the brain are differently analysing that information. So, I am breaking down the whole image into pieces. What connected that? I don't see only colour or position, I see the object. We still do not know the answer. There are 40 areas of the brain at least which are involved in the simple task of visual functioning, the most studied.

Consider a cup of coffee. You may think that a cup of coffee is only an object, that it has length, breadth, shape, size and colour. No. A cup of coffee arouses many other sensations in you. It arouses the sensation of smell, warmth, the shape, even the sense that this coffee comes out from the coffee beans and as I have often said it might also remind you of the company that you had when you had a special cup of coffee. That moment of time might be the most pleasant moment of your life which you recall with all its intensity and emotion combined together. So, it is not just a cup of coffee and each component, as you see, involves different parts of the brain. Therefore, the question is what brings it together.

The matter doesn't end with this. Daily routine is the mundane aspect of our existence. Sri Ramakrishna said, ". . . in the depths of meditation there arises the consciousness of the impartite, the indivisible Reality. Body is forgotten, mind ceases to function -there remains pure consciousness". Shri Ramakrishna is not alone. Many saints in diverse cultures have come to similar conclusion. All that I have said so far has nothing to approach the answer to this question as of today. The very eminent, distinguished physicist Erwin Schrödinger said that the whole world is a project of the one undivided cosmic consciousness and each of us is an integral part of it. These are philosophical questions dealt with by scientists and non-scientists equally but still not within our ability to grasp them in hard core data, as Prof. Sudarshan mentioned earlier in his speech in beautiful language.

I close by saying that my conclusion remains unchanged after 10 years. There are enough unexplained observations which need to be resolved before the answer to the internal question of brain,

mind and consciousness can be found. No doubt this question is difficult but the past could be made easier and better illuminated if the neuro-biologists take along with them the natural scientists, the psychologists and the philosophers on this journey. I hope this conference will help in this direction.

REFERENCES

Campbell, A. 1980. *Seven States of Consciousness.* London: Victor Gollancz Ltd.

James, W. 1892. *Textbook on Psychology.* London: Macmillan.

Kothari, D.S. 1980. *Atom and Self* (Meghnad Saha Lecture 1978). In *Proceedings of Indian National Science Academy,* 46A, pp.1-28.

Penfield, W.P. 1975. *The Mystery of Mind: A Critical Study of Consciousness and the Human Brain.* Princeton, N.J.: Princeton University Press.

Popper, K.R., and J.C. Eccles. 1977. *The Self and Its Brain.* Berlin: Springer Verlag.

Sperry, R.W. 1966. *Mind , Brain, and Humanist Value. Bull Atomic Science* 22, pp.2-6.

Tandon, P.N. 1993. *Consciousness: Clinical and Beyond* (Presidential Address, INSA, December 30, 1992, New Delhi, India). In Procedings of Indian National Science Academy, B 59, pp. 5-51. Reproduced by Indian Institute of Advanced Study, Shimla, 2001, under their *Dissemination of Knowledge* Series.

Tandon, P.N.1999. *Exploring Consciousness! Neurobiological Approaches* in *Scientific and Philosophical Studies on Consciousness.* Ed. by S. Menon, M.G. Narasimhan, A.Sinha & B V Sreekantan. NIAS: Bangalore, India.

Tandon, P.N. 2000. *Theme Address-Round table Discussion on Consciosuness.* In *Proceedings of the Round Table on Consciousness,* Oct 12-13, 2000. National Brain Research Centre Symposia Series-1, National Brain Research Centre, Guragaon, India, pp.17-22.

Tandon P.N. 1995. *Concepts of Consciousness: Brain-mind Relationship.* In *Decade of Brain India/USA Research in Mental Health.* Ed. by S.B.Kaslow, S.Srinivasa Murthy, C.V.Coelho, pp17-22.

Wald,G. 1979. *Life and Mind in the Universe.* In The Evolution of Consciousness. Ed by K.Gandhi, New Delhi: National Publishing House, pp.1-16.

Alternate Paradigms of Knowledge

KAPILA VATSYAYAN

I am grateful to the Indian Institute of Technology, Kharagpur, Department of Humanities and Social Sciences, particularly Dr. Chhanda Chakraborti, and the Indian Institute for Advanced Study, Shimla, for having invited me to participate in this very important international conference on mind and consciousness. I accepted the offer to participate because I was certain that this occasion would be a great learning experience for me. Despite the fact that I am neither a scientist nor a technologist, nor a philosopher nor a linguist, I have been actively conscious of the interpenetration of the diverse domains of knowledge and their mutual dialogue. As one concerned with artistic experience, expression and communication, it has become clear to me that artistic creation and its critical discourse is effected by other domains of knowledge and disciplines and in turn directly or indirectly affects other domains. Both the sciences and the humanities employ metaphors to communicate. This, to me, is an artistic act. Although the methodological tools of each discipline are distinct but in the last analysis, the attempt is a search for essence and for truth. Each attempts to create or recreate or re-articulate a whole, be it the scientist, philosopher, poet or artist. Long long ago the Ṛg Veda has said truth is one but the wise call it by different names. Truth is one, the paths are many.

A few years back I happened to visit Visva Bharati for a symposium on the dialogue between Tagore and Einstein. It was at this symposium that I heard the views of many eminent scientists on Einstein's view and the developments in scientific theory post-Einstein. These articulations were juxtaposed with the poet Rabindranath Tagore's views on creativity and experience. I was reminded of the several discussions on aesthetic theories in the Indian tradition. To me it appeared that after an arduous journey in different directions, there was a confluence which was taking place between the ancient and the most modern, between the intuitive, nondiscursive mode and the quantifiable measured and rational.

My spontaneous and equally unmeasured response to the discussions resulted in this gracious invitation, thanks to Prof Abrahim Khan of the Trinity School Cambridge. I am happy to be in this gathering which will approach the issue of mind and consciousness from many different points of view.

Inadequate as I may be and an outsider to the discourse on the nature of mind and consciousness in the domains of the physical, natural and neurosciences, it appears to be that there is some relevance of speaking about systems of knowledge which also addressed the issue but not through empirical experimentation. Theirs was the route of speculation and experience. They began by subscribing first and foremost to the principle of 'essence', call it a state of formlessness, intangible but comprehensible by the mind of man. Interconnectivity between the parts and the whole was a basic tenet, as was the consciousness of a continuous process of transmutability. Matter could be transformed into energy and vice-versa. From these primary perceptions there evolved many different streams. Basic to the plurality of views was the awareness that the universe could only be comprehended by the human mind as a part of the universe and not an objective observer of a phenomenon outside him.

In this context, the conversation between Tagore and Einstein assumes a renewed importance. Undoubtedly Einstein was deeply concerned with two conceptions of the universe. He asked if Tagore believed that the world (a) was a unity dependent on humanity, (b) was a reality independent of the human being. Tagore's reply was unequivocal. "When our universe is in harmony with man, the eternal, we know it as truth. We feel it is beauty". Einstein responded, "this is a purely human conception of the universe", and Tagore almost shot back "there can be no other conception. This world is a human world – the scientific view of it is the view of the scientific man. Therefore, the world apart from us does not exist; it is a relative world depending for its reality upon our consciousness."

Einstein continued to try and assert the existence of a reality beyond human perception. However, he conceded "I cannot prove that scientific truth: it must be conceived as a truth that is valid independent of humanity, but I believe in it firmly", and said "the initial problem is whether truth is independent of our consciousness".

The conversation took place on July 14, 1930 at Einstein's residence in Germany. The issues raised have engaged generations

of post-Einstein scientists and philosophers. It is not necessary for me to list them at this gathering. Ilya Prigogine said in an interview with Ranjit Nair not long ago sometime in 1995 "Well you know there was something in which Tagore was right and Einstein was following without knowing it a mistaken dualistic approach. Einstein says that it is a miracle that the universe is comprehensible. That this is a miracle, is a dualistic view because then you separate man from the universe. If man is the outcome of the universe, if he is a part of it or an expression of it, then it is not astonishing. This is the holistic point of view that is more common to the Indian tradition and which Tagore expresses. Einstein, unlike Tagore, very much separated man and nature".

These brief excerpts are indicative of the fact that over a long period, in many different ways, there have been two distinctly different approaches to the universe and man. Also inherent in the conversation and elsewhere is the concern with verifiability and proof. An Einstein cannot prove that there is a universe independent of man but he believes in it; Tagore believes that the truth lies in the experience of Man and the gradual but sure process of transcendency of the individual Mind to a universal consciousness. He elaborates upon it in his book *Religion of Man*.

The scientific discourse of the last few decades from differing points of view, and diverse methods of exploration, has revolved around the crucial questions of matter and non-matter. By breaking down matter into the smallest units, there was the entire discussion on wave-particle, which moved beyond space-time continuums. To the fore came the more metaphysical questions of comprehending the universe by the human, and the very basis of perception and the act of perception. Questions of uncertainty and certitude arose as also those of objectively and subjectively and finally of the models of the universe as closed or open systems. Each of these has also led us to ask the question of absolute and relative categories and most of all about the acceptance or rejection of the level of the immeasurable and measurable.

Obviously this is in deep contrast to all that is understood and known by Newtonian determinism, Cartesian dualism and Freudian fragmentation and its fallout in 'modern intellection and its structures'.

It is against this background and with some hesitancy and acute consciousness of inadequacy that an attempt is being made here to

draw attention to some fundamentals or one may say overarching approaches to these issues and the consequent evolution of a knowledge system which for facility of understanding we have called an alternate knowledge system. In fact one could substitute the word 'alternate' with the word 'primary', original and indigenous to most part of the world, pre-European Renaissance and certainly pre-Newtonian science and Cartesian dualism.

Despite the diversity and the seemingly bewildering complexity of the several Indian systems of thought and the rhythm of the changes in interpretation and re-interpretation, there continue to be some key notions of the universe and man, of space, time, primal elements which permeate the diverse perceptions. The perceptions may be called in general terms worldviews or views of the universe, which constitute the foundations of a knowledge system or knowledge systems.

In the case of science, it was the search for both the nature of matter and its breaking which led to a sense of mystery and acceptance of ambiguity and uncertainty. It was also the process of perception, which led us to define the nature of systems, predictability and unpredictability, chaos and order. The other or alternate world began by accepting almost as an article of faith, that there was a realm beyond the knowable and measurable, call it an intuitive insight or experience. The micro-mind of man was conscious of an immensity and extensiveness. The consciousness of the unknown and the immeasurable reverberates throughout. The first articulation in the *Ṛg Veda* called the creation hymn lays the foundation. The *Nasadiya Sukta* –Rv. X. 129 begins by accepting "At first there was neither Being nor non-Being. There was not air nor yet sky beyond. In the beginning consciousness arose which was the primal germ cell (*bija*) of the mind, seers searching in their hearts with wisdom discovered the connection between the Being and non-Being".

Apart from this great creation hymn, there are others which revolve around the three central concepts of *tapas* (ardor) *ṛta* (order) and *satya* (truth). These are pursued in many streams constantly over a long historical period.

The consciousness of *tapas, agni* and the sun as the seed of *agni* (energy) as *agnibija* engage the mind of the Ṛg Vedic thinker. *Tapas* is the proto-cosmic energy which is the double principle of the whole of reality. *Ṛta* (order) is the structure, the first principle, the contexture of reality, and *satya* (truth) the substance and material

principle, the concrete and crystallized. However, none of these are static. If *tapas, agni* is the energizing principle, *ṛta* (order) is the organizing principle. The world is not a chaos, it is a cosmos, not an anarchic mass. An organizing principle facilitates diverse configuration of harmony. *Satya* (truth) is the substantive principle which provides consistency through a variety of paths.

In each of these concepts, it is clear that there is the acceptance of the notion of that which is unseen but effective. The sun is known by its multiple rays, but not directly perceived. Similarly, *ṛta* (order) is perceived indirectly by the constant rhythm of renewed equilibriums and satya (truth) by a constant search for the essence.

Naturally it is in the mind of man that there is the consciousness of the vast fathomless universe, the galaxies and the sun, the processes of order and chaos. This consciousness evolves into a fundamental investigation into the relationship of the macro-micro, of the relationship between the vast infinite and the tiniest finite. In traditional terminology we recognize it instantly as the discourse on the *pinda* and *brahmanda.*

The consciousness of the macro and the micro, and the constant rhythm of new configurations of order (*ṛta*) lead to the principle of reversibility and the interchangeability of matter and energy – popularly called *jada* and *cetana* translated generally as inanimate and animate. This consciousness of the relationship of macro-micro, of matter and energy naturally gives rise to very different notions of space, time and the primal elements. Each concept is explored at its macro-micro levels, and its grossest material level and its subtlest non-material level. Both the telescope and microscope of the mind are employed.

Man's consciousness also recognizes that exactly as in the first primordial moment, so also in thought and speculation, there is an experiencible but immeasurable level. This should not be unfamiliar to those concerned with the recent active discourse on the difference between the mathematical and computable, not to speak of the recent debate on the limits of mathematics itself.

The knowledge systems we shall view evolve from these perceptions. They begin from the premise that there is a category of knowledge which is experiencible, but not computable. It is not beyond comprehension in the consciousness of man's mind, but it is beyond measure and location. Modern science attempted to measure the macro and micro, energy and matter by breaking each

component to its tiniest units, it reached a point where there was not further scope for measure and computation. There was a sense of uncertainty, ambiguity and even mystery. In the other or alternate knowledge system, the mind of man began by accepting the mystery, the ambiguity and the immeasurability. In short the knowledge system began by accepting a level of undifferentiation, a wholeness, a single unified field. This level of knowledge, more of experience and understanding was the 'beyond' termed as *para vidya*. The *Ṛig Veda* had spoken of the processes of the mind which moved from the state of the unformed, to form and multiple forms, and to a consciousness of beyond form. A sophisticated discourse follows from the adage of the *arupa, rupa, pratirupa* to *pare rupa*. The *para vidya* is analogous to *paré rupa*.

The distinction between the two levels of knowledge, i.e. *parā-vidyā* and *aparā-vidyā* must not be equated to sacred and profane levels as has been often done. It must be remembered that *parā-vidyā* is also termed as *brahmavidyā* or *ātmavidya*. Both terms allude to the consciousness at the macro and micro levels of comprehension which is immeasurable. The emphasis is on the experiencible, the comprehensible but not measurable. It is not 'sacred' as a religious category. Since this knowledge is naturally not quantifiable, it is both immense and intense, extensively dissipative as also internal and introspective. Its verifiability is in its authenticity of experience. It is almost as an article of faith as Einstein said - that "I know that a universe exists".

The other level of knowledge is the *aparā-vidyā*. This level of knowledge is subject to differentiation and classification. Many classificatory of systems are evolved. However, each of the systems accepts the intrinsic relationship of the macro-micro – the material, non-material, and the constant possibility of reversibility. Further coordinates are established with different aspects of the phenomenal world. Different streams of knowledge are recognized, which we term as disciplines, and there are the broad classification of the knowledge into different domains, goals and values of human life. A variety of *Sāstras* emerge; each has its own system of verifiability (*pramāna*) and tools of investigation. Although each of these investigative intellectual tools are distinct and diverse, there is an inbuilt system of connectivity between and amongst the disciplines. None is a closed system, there is interpenetration between and amongst them.

Our concern here is not with the classification of disciplines, e.g. *vyākarana, mimānsa, nyāya, jyotisa, gaṇita, āyurveda, itihāsa, purāṇa* or with *sāstra* and *śilpa* nor with the division of knowledge into *triya, varta*, etc. Instead, it is with the nature and manner of classification, which makes it possible to explore some key concepts in each of the disciplines and to make links across the disciplines. Any discerning reader of the primary texts in Sanskrit, Pali, Prakrit is struck by the fact that certain key concepts are investigated systematically in the disciplines of *āyurveda, jyotiśa, gaṇita, darśana*, as also the arts *śilpa, nātya, sangita*. They are systematically investigated in the *śāstra* and govern *prayoga*. Further, each discipline takes into cognizance discourse in the other disciplines. The discussion thus is not only interdisciplinary in character, but also one in which interpenetration takes place of one discipline into the other. If the texts of music discuss the body-system in detail and give considerable attention to mathematical equations, the *Āyurveda* texts situate themselves in philosophic systems.

In such a pervasive situation it becomes necessary to identify some key concepts and terms, which permeate most of the disciplines and to follow the contours of the discourse in and around them in a few disciplines. It was this exploration, which enabled one to identify the methodologies adopted in the knowledge systems for evolving categories. It was clear that no category was absolute or fixed, it was also clear that the categories were also not insulated or in confrontation with each other. The principle of complementarity rather than opposition was fundamental. Also there was no absolute progression or hierarchy. Dominant and subordinate were in context, space and time could be reversed. It will be obvious that such an approach is in distinct contrast to that other system which is based on viewing categories as binary opposites, or one in which conflict and resolution are the inevitable sequential phases. Here complimentarity and even concurrency are evident. Multi-layering and multi-texturing of categories and interpretation was natural.

We have spoken at the very outset of the acceptance of a level of knowledge, which is experiencible but indefinable. We may also allude to the recognition of a glacial level of thought as a unified field from which multiplicity flowed. In the matter of concepts and key notions, it is the first and foremost concern with the whole. The one which is indivisible.

The *Ṛg Veda* had spoken of *Ekam sad viprā bahudhā vadanti:* others,

through a variety of disciplines, speculative and empirical, are equally concerned with the notion of one – the complete. The *advaita* that we all know is not non-dual, it is adual. The distinction is an important one. It is not necessary to dwell on the issue of the one and the many here. Suffice it to say that the knowledge systems first explore the many paths to investigate the one. The philosophic system of the Jaina *Anekānta* is one outstanding example. There are others. The discussion on *ātman* and *brahman* in the tradition has to be viewed in our context of the concern of the system with investigating 'oneness' or 'wholeness', or centrality. *Ātman* and *Brahman* are terms of reference of centrality at the micro and macro levels. No more elucidation need be given here of these terms. Perhaps two excerpts from the *Mundaka Upanishad* would be relevant:

> That which is invisible, ungraspable without family without caste (*avarna*) without sight or hearing. It is without hand or foot, eternal all permeating, omnipresent, exceedingly subtle that is the Imperishable which the wise perceive as the source of being.
>
> As a spider emits and draws in (its thread). As hawks arise on the earth
>
> As the hairs of the head and body from a living person
>
> So from the imperishable arises everything here.

Mundaka 1. 6-7

We can now move on to the concepts which appear as pairs. They are not dualities in opposition. There are countless pairs which permeate discussions in practically all disciplines and discourses. In popular parlance we are familiar with the terms *sukha-dukha* (happiness-suffering) *sita-usna* (cold-hot), etc, etc. The question to be asked is whether they are in opposition as absolute categories, or in a relationship of complementarity. The answer is clear that while they are distinct, each is complementary to the other; and can be in a sequential cycle as day night (*usas-rātri*), etc.

For our purpose here the most important example of the pairs is a relationship of complimentarity is the image of the *bhoktā* and the *drstā* in the *Upanishad*. The experiencer (eater, *bhoktā*) and the observer (*draśtā*, seer) are two birds on the same branch. One cannot survive without the other. The verse of the *Rg Veda* 1.164 is repeated in *Katha Upanishad*, *Mundaka Upanishad* (3.1.1) and *Svetsvatara Upanishad* sums it up:

> Two birds, fast bound companions clasp close the self same tree, Of these two, the one eats fruits, the other looks on without eating.

But neither can exist without the other. The concept of the *bhokta* and *draśtā* is a remote anticipator of the contemporary discourse on the observer and the observed. With a little stretch of imagination it almost anticipates the problem of Schrödinger's cat, but of that we shall say no more. Here we are concerned with the classification of dualities, which are interdependent and not exclusive. In traditional language they are called *mithuna* and not polarities. Each time the pairs are set up, the questions asked are whether they are insulative, combative, or complementary. Also the questions asked are whether it is the one which breaks into two, or it is two becoming one. The answer is always that it is the one becoming two or many and not the aggregation of the two or many into one. The *ardhanārisvara* is the most concrete visual symbol of the concept.

We address next the methodology of evolving categories in triads rather than pairs. The triads are numerous and cover a vast range of disciplines. The *triya* division of Vedas and knowledge is well known. Here each concept is explored at its material, non-material and internal experiencable dimension. Basic is the division of considering the concept at its mundane terrestrial level i.e. at the material level (*bhautika*), and then at its transcendental or celestial level (*daivika*) and then at the experiential level of the self, the introspective inner level (*ādhyātmika*). The three levels provide the possibility of a multidimensional understanding of the concept in space, time and inner deep experience, as consciousness.

Important amongst the triads is the concept of the *triloka* and *trikāla*. Both space and time have one dimension, which is the immeasurable and indefinabale and unquantifiable. Space is both vacuity and fullness. It is *kha*, *vyoma* and *ākāśa*. Much has been written on the subject from A. K. Coomaraswamy to Halbfass in recent times. This dimension of infinite space is one level. However in the world of differentiation there is the notion of *loka*. Here too there is first the division between *aloka* and *loka*. The latter is the sphere of direct perception. It is both space and what fills space. It is both the object of perception and the act of perception. It is the measure, as also the act of measurement. Although there is great complexity of discourse on

the term, specially in Jaina philosophy, at one level it alludes to the division of the universe into three spheres. This is *triloka*. There is the *bhuloka* (the earth), the *bhuvarloka* or *antariksha* (mid-space) and *svarloka* or *vayuloka* (as the region beyond). In Vedic ritual there is final oblation to *bhu bhuvah svah*. The three are interconnected. The subject of space (*ākāsa*) is vast and complex. Even I have edited a whole volume on it. Here we have barely touched upon one aspect of the subject.

Similarly 'time' too has an indefinable dimension. It is immeasurable. At another level it is measurable as the expansive time of *mahākāla*, the macro or the micro, *nimesa* (i.e. blinking of an eyelid). It can be measured as aeons, as in Indian astronomy and is subject to computation and in terms of the micro units of *kshana* (moment). Its broad divisions as in the case of *triloka* are *trikāla*. Each phase of the *trikāla* is related to the other, as in the case of *triloka*. The conception of the *kālachakra* alludes to the attributes of motion and movement. It is not linear progression. Indeed each constituent is in a relationship of cyclicity with the other. There is the perception of coiling and recoiling. Linearity is subsumed in the cyclicity. Little wonder that the coiled serpent, the snake biting its tail, the continuous line of eternity, the wheel (*chakra*) with its felly and spokes becomes a pervasive common symbol of the conceptions of time discussed in the disciplines of philosophy – mathematics, astronomy and medicine and, of course, the arts.

The differences between these notions of space and time and those of knowledge systems, which evolved in post-renaissance Europe and crystallized as the scientific principles of the irreversibility of time are fundamental. They have far reaching implications in all domains of knowledge and perception and, of course, creativity specially in the artistic traditions of India.

However, to move on from the very important questions of space, time to another series of triads, let us take as an example the category of *gunas*. The *gunas*, along with the two other related categories of *dravya* and *dhātu* pervade discussions in practically all disciplines of the knowledge system. In popular discourse we refer to the three *gunas*, the *tamas*, *rajas* and *sattva* as being equated to the inert, active and the purified state of matter and being. Also the term is used for multiplication and replication. A slightly deeper attention will reveal the nature of the concept, and the division into three. In the mutual relationship of the constituents there is possibility of reversibility.

Here also first and foremost is the acceptance of a state or stage which is without attributes. This is the formless, it is the nascent. At the level of the tangible in its primary meaning it is conceived as strands of thread which are entwined as a rope. As is well known the *gunas* are frequently discussed in the *Upanishads*. They constitute the bedrock of *Darsanas* specially *Sāṅkhya*. They are central to the system of medicine in *ayurveda*, are a core mathematical principle. They are alluded to in early chemistry and occupy an important place in discussions on theories of aesthetics and govern artistic praxis.

Buddhist philosophy discusses them and the *Gita* devotes considerable attention to them. Our purpose here is not to dwell on the complexity of the discussions in each of the streams of thought and philosophic schools and *ayurvedic* texts. Instead it is to again ask the question whether the division of the category into a triad, views each component of the triad as absolute entity and whether the principle of reversibility applies here or not. Also it is to ask the question whether the category of *guna* is a material object or whether the category alludes to process.

A perusal of a vast body of texts in different disciplines makes it clear that despite the complexity and multiplicity of meaning and interpretation the category of *gunas* invariably embody the attribute of strands of thread entwining and folds of a single continuum such as cloth or robe. Further as in the case of *loka* and *kāla*, the emphasis is on process and the potential of transformation of one component into the other. The relationship is of complementarity and reversibility. They are not absolute categories.

Thus, the inert dark matter *tamas* can be ignited to activity and become *rajas*. The *rajas* in turn can be further refined and purified to become *sattva*. A reverse process is possible. Inherent is the idea of latency and potency as also the unmanifest and manifest.

Whether in *āyurveda* or *Sāṅkhya* there is the acceptance of a primary state of the unmanifest (*avyakta*). The latent (unmanifest) supreme nature (*prakriti*) is the progenitor of all created things, says *Susruta*. She is a self-begotten and connotes the three qualities of *sattva*, *rajas* and *tamas*. The *ayurveda* texts also speak of the connection of the three *dosas* namely *vata*, *pitta* and *kapha* with the three *gunas*.

The biological and psychical are both cause and effect. The term or category as is well known assumes great philosophic significance

in the *Sāṅkhya*. The *Sāṅkhya-Karika* (13) gives the simile of a flame to explain that although the three *gunas* are distinct from each other, together they serve the same purpose as the three entities of wick, oil and fire in a flame. There is a very detailed and sophisticated discourse on the *avyakta* (unmanifested, latent) aspect and the *vyakta* (manifested, potent) aspect. For us it is important to note that the discussion on this category also takes into cognizance the undifferentiated and the differentiated, as also the mutuality of the components of the different entities.

The category of *guna* is closely linked to the categories of *dravya* (substance) and *dhatu*. *Dravya* derived from the root *dru* alludes to the essential quality of fluidity of melting and dissolving. In turn *dravya* becomes the process of liquidification of the *gunas*. *Dravya* is also material and non-material. While the detailed discourse on *dravya* in the schools of philosophy particularly as elucidated in the commentaries of *Vyasa* on the *Yoga-sutra* of *Patanjali* or in *Vaisesika-sutra* need not detain us here, it is important to note that the debate revolves around the central issue of substance perception, the act of perception and the identification of qualities and attributes.

Implied in the vast discussion on the categories of *gunas*, *dravya* and *dhatu* is the intrinsic relationship of the material and non-material, the processes of liquidity and inter-relationship, as also the aspect of primary and secondary specially in respect of dhatu. While each category has sub-divisions of three, seven, etc., they also interpenetrate into each other in many meaningful ways.

Logically the triad of the *gunas*, and the cluster of *guna*, *dravya* and *dhatu* is closely related to and even dependant on another triad, which we recognize as *sthula* (gross), *sukshma* and *para* (beyond). While liquidity and fluidity was implied in the definitions of *tamas rajas* and *sattva* and was explicitly stated in the case of *dravya*, the conception of *sthula*, *sukshma* and *para* gives almost a paradigmical model of considering the process of a transformation of the world of measure to the measureless, from the solid to the liquid, to the gaseous or beyond. Naturally *sthula* becomes a cognate of *tamas*, and *sukshma* of *sattva*. They are not identical but are comparable. Equally important is the relationship of the concept of *sthula*, *sukshma* and *para* to *avyakta* and *vyakta* the unmanifested and manifested. Most important is the fact that this triad now refers to the nature of knowledge. Knowledge itself is graded into the gross and subtle and that which is beyond discursive thought. A gradual movement from the gross to the subtle and the intuitive is implied. The *Bhagavadgita* expounds the concept of the three interconnected terms *sthula*, *sukshma* and *para* as it does on the *gunas*. At a gross level knowledge manifests itself in different forms – the primal elements each, water, fire, air, ether, mind, intellect and consciousness of ego. But there is another subtler form of knowledge of the *atman*, this is supreme knowledge, and finally there is that experience of the cosmic, which is termed as *param-purusha*. The *Katha Upanishad* also gives a similar grading of the gradual movement from the gross to the subtle – from the body to mind, to intellect to a higher consciousness.

> Higher than the senses are the objects of sense
> Higher than the objects of sense – the mind (*manas*)
> And higher than the mind is the intellect (*buddhi*)
> Higher than the intellect is the great self (*atman*)
> And further
> Higher than the great is the unmanifest (*avyakta*)
> Higher than the unmanifest is the person (*purus e parah*)
> Higher than the person there is nothing at all .
> (*Kathopanishad* III.11)

Bhartrihari elucidates briefly on the levels of knowledge and the instrumentalities. He concedes that perception and the act of perception is primary, the level of that perception and the consciousness of the act of perception determine its gross or subtle nature. Practice, *abhyasa* the continual practice of discrimination is essential. The rudimentary level of sense perceptions is the beginning, the gradual refinement is the journey to refinement and subtlety, the sudden illumination of the totality and wholeness is the beyond and *para*.

The discussion on the levels of gross, subtle and beyond or transcendental is extended to the primal elements, the sense organs and sense perceptions. The schools of philosophy – Tantra and in all streams of thought Hindu, Buddhist and Jain, discuss the levels of the gross and subtle at great length. The identification of the gross and subtle levels of matter and cognition is of great relevance to us in the light of the recent discussion in the neuro-sciences on the levels of lower and higher consciousness.

We have now to move on to the category of concepts, which may be roughly termed as the five folds. In the earlier group the divisions were broadly as three and triads were formed. In another grouping

there is the division into five, at the micro and macro-level. The human is first and foremost aware of the five-sense organs and the five sense perceptions. This is basic and elementary. As obvious is the awareness of the primal elements – earth, water, air, fire and space. If the first is basic to body of the biological man, the second permeates the universe and man's consciousness. The worldview and the knowledge system acknowledge the relationship and interpenetration of the elements and the body system, as also the possibility of the micro body to expand and be the Cosmic Man. A mighty structure of concepts is evolved on the basis of the relationship of the micro-macro of the small man (*purusa*) and the Cosmic Man (*Purusa*). The paradigm is the body of Man with the sense organs and sense perceptions as vital constituents. Naturally a discussion on the Mind of Man follows. It is not necessary to mention the *Purusa Sukta* as a paradigonical model of the universe. This is a first powerful articulation of interlocked structures.

The concept and category of *indriya* at the micro level expands to macro-level and is integral to the comprehension and its articulation on the primal elements – first three and then five. These discussions bring to fore the fact that in this worldview and the knowledge systems, there is the cognisance of the intrinsic wholeness. There is neither dichotomy nor fragmentation of the senses, body, mind, intellect and consciousness. While the body can be identified distinctly as chariot, it is neither inanimate nor separate from the mind, and naturally mind from consciousness. This is the message of the *Katha Upanishad* where it speaks of the body as the chariot, and the senses as the horses, and mind as the reins, and intellect as the charioteer. The sense organs and sense perceptions are indispensable as horses are to a chariot. The discourse on the *indriyas* is as complex and sophisticated as that on the *gunas*. Our concern once again is not to outline the development of the concept in various disciplines and branches of knowledge. It is instead to point at the approach to the sense organs and sense perceptions and to point at the interdependence which is established between and amongst them and finally to underline the systems of coordinates and correspondence which is developed.

The *indriyas* are central to the discourse at the most physical and metaphysical levels, micro and macro, outer and inner. *Indriya* derived from *Indra* (self, *atman*) alludes first and foremost to vigour, resplendence and energy. The term is also related to *retas* (seminal fluid). The concept refers as much to the physical – biological sense organs as also to the non-material attributes of prime seminal energy, vigour and resplendence.

The *Atharva Veda* explicitly speaks of the five sense organs, and mind as the sixth.

"These five sense organs with the mind as the sixth
Within my heart inspired by Brahman by which the awe inspiring is created
Through them to us be peace."

(A.V. XIX 9.5)

The *Upaniṣads* are replete with a discussion on the sense organs and sense perceptions. Each Upanishad focuses on one or the other aspect of the sense organs and sense perceptions and the mind. Intrinsic to the discussion is their interdependence. Both in the *Chandogya Upaniṣad* and the *Brihadaranyaka Upaniṣad* occurs a most memorable dialogue amongst the sense organs. Each tries to assert its superiority. In the end, it is clear that without *prāna*, the vital breath, they are inoperative and in the absence of any one the others are disabled.

The *Katha Upaniṣad* speaks of them as the indispensable horses of a total system comprising the senses, body, mind, intellect and consciousness. We have referred to it earlier. There is then the discourse on the *indriya*, which have an inward and outward movement (*antar* and *bahya*).

Most important of all is the dimension of the micro and the macro, the human and the cosmic. The *Ṛg Veda* articulates it powerfully.

The Moon was born from his mind,
the sun came into being from his eye,
from his mouth came Indra and Agni,
while from his breath the wind was born.

From his navel issued the air,
From the head unfurled the sky,
The Earth from his feet, from the ear the four directions.
Thus have the worlds been organized.

(R.V.X./ 90 – 13-14)

Already a system of correspondence is established between the primal elements and the sense organs.

The *Mundaka Upanishad* echoes the *Ṛg Veda* conception of the cosmic Man, the senses are part of His Being :

From Him is produced breath (*prāna*)
Mind (*manas*) and all the senses (*indriya*),
Space (*kha*) wind light water
And earth the supporter of all

(*Mundaka Upanishad* II, 3)

This relationship of the micro and Cosmic Man, of the sense perceptions and the primal elements runs through like a sustained thread in practically all disciplines and branches of knowledge, ranging from *ayurveda* to the schools of philosophy (*darsana*) and the arts. The theme of the relationship of the senses with each other, their interdependence and their correspondence with the primal elements is systematically pursued. The system of correspondence between the sense organs and perceptions and the primal elements (*bhutas*) is near universal.

For example, Charaka says "Among the existent sense organs which are comprehensible by inference and are composed of combination of the resultants of five *mahabhutas* the visual auditory, olfactory gustatory and tactile organs have predominance of *tejas, akasa, prithvi, apa* and *vayu* respectively. Thus the sense organs perceive the respective objects according to predominance of their constituent *mahabhutas* because of their similarity of nature and specific capability." (*Charaka Samhita* VIII. 14).

Conversely the philosophic schools discuss in details the relationship of the sense organs, sense perceptions the mind and consciousness *Sāṅkhya* develops a rigorous system of dividing the *indriyas* into *karmendriyas, Jnanendriyas, manas* and the five *tanmātras*. The subtle *suksma tanmātras* are related to the five gross elements (*sthula-bhutas*) ether, air, fire, water and earth.

While *Sāṅkhya* and *Sāṅkhya-Yoga* discuss the subject of the *indriya* in great detail, and a psychosomatic systems developed of equal importance is the discussion in *Nyaya*. As we know *Nyaya* recognizes four sources of true knowledge viz. perception (*pratyaksa*), inference (*anumāna*), analogy (*upamāna*) and credible testimony (*sabda*). In the contest of the first, i.e. perception (*pratyaksha*) there is a detailed discussion on the self, the body, the senses and their objects, cognition and the mind. The discussion, the questions and answers

given reveal the concern of the system to establish inter-relationships and correspondence. According to one view each sense organ in composed primarily of the stuff it grasps, thus the eye is made primarily of light, the ear of *ākāśa*, etc. Thus a one-to-one correspondence is established, e.g.

Elements	*Sense qualities*
Earth	smell
Water	taste
Fire or light	colour
Air	touch
Ākāśa	sound

It will be clear from the above that the diverse schools are exploring many paths of establishing a series of correspondences. These have to be understood from within the context of the particular school or discourse. However, the preoccupation with establishing correspondences, is a recurring feature of the many systems.

The concept of the *indriyas* also intensely engages the Buddhist and Jaina philosophy and the arts. The discourse has to be comprehended from within the overall worldview and the philosophic principles. *Abhidharma kosa* devotes a full chapter to the subject (Chapter II). While it would be difficult to summarise a complex argument, it is perhaps relevant to mention that the *Abhidharma kosa* considers the *indriyas* to be twenty two. Of these five belong to the realm of matter and form (*rupa-skandha* and *rupa dhatu*) along with two others namely the sex organs. Others are called *jivitendriya* (vital), they are considered to be *bhautika*, i.e. related to the primal elements. This group interacts with consciousness (*vijnana*) and broadly belongs to the substratum of consciousness and mind (*citta*). The first seven are considered external or outward (*bahyaka*) and others are internal (*ābhyantaraka*).

This simplistically stated summary would make it evident that the *Abhidharma kosa* is also concerned with the sense organs and sense perceptions, their inter-relationship and their relationship to the primal elements. We cannot pause here to dwell upon the atomic make-up of the five *indriyas* discussed at great length in the *Abhidharma kosa*. Also we have to omit the important discourse on the *indriyas* in the schools of Jaina philosophy.

Nevertheless, even this brief enumeration is sufficient proof of the fact that the sense organs, sense perception and the act of perception occupy a central position in the knowledge system. All levels are systematic pursued, there is no negation of the senses and sense perceptions or the body. Instead the human is an integrated system of the senses, body, mind and consciousness. Processing and a continual yoking together and refinement are the key concerns. Establishment of correspondences is characteristic. A concept enlarges from its core as physical material to ever widening circle of meanings and connections.

The *indriyas* and the primal elements interpenetrate. We have already referred to the *Ṛg Vedic* hymn and its echoes in the *Upaniṣads* and other texts. The group of five elements, (initially three) are another cluster which permeate discussions in all disciplines and branches of knowledge. Naturally the elements are primal. Little need be said. All civilizations and cultures have reflected upon them. In some culture and civilizations there are three in others four and in most others five. Within the system we are speaking about, there is a marked preoccupation with them from the earliest times. Again the Vedas, Rg Veda specially is full of references to the elements. In many hymns the sun is invoked as agni, fire, he is the fire-seed (*agnibija*). The *Bhumi Sukta* of *Atharva Veda* invokes the earth. There are countless other hymns devoted to *varuna* (water), *vayu* (air) and *akasa* (space). The corpus is extensive and the discourse varied and multifaceted. An ecological system is developed.

The *Upaniṣads* dwell upon the elements, as they had done on the sense organs and sense perceptions. The schools of philosophy are engaged in a lively discourse on them, and the arts concretize the conceptions through theories and actual praxis be it architecture, sculpture, music or dance. The consciousness of the primal elements and their interaction is pervasive.

Although speculation on the general structure of the universe is articulated in the Vedas, it is in the *Upaniṣads* that we find a clear formulation of the entities and their function. For example the *Taittiriya Upaniṣad* states:

> Now with regard to the world, the earth is the prior form, the heaven the latter form. Space is their conjunction. Wind the connection. Thus with regard to the world.
>
> Now, with regard to the luminaries. Fire is the prior form, the sun the latter form. Water is their conjunction, lightening the connection. Thus with regard to the luminaries.
>
> (*Taittiriya Upaniṣad* 1.3)

Sankhayana Aranyaka elucidates the concept of the five *mahabhutas* which combined with other minor entities form the *Sarvabhuta Samhita*.

> There are five vital elements. These are earth, wind, space, water and fire. These are the combination jointly or together.
>
> *Sankhāyana Āranyaka* VIII. 22.

These articulations are early examples of creating a general concept of matter. It is important to note that although of the entities four or five are identified, there is always the consciousness of their cojointedness and interpenetration.

The schools of philosophy, each with their distinctive methodologies, dwell upon the entities and their relationship. The transformative view of *Sāṅkhya* for example conceives of an original eternal non-differentiated matter called *prakṛti* (literally base). This is then transformed into different constituents. Further equivalence is established between the micro subtle *tanmatras* and the five elements. A system of correspondence is established between the subtle and gross, the micro and macro. In time to come, the *tantras* some basing themselves on *Sāṅkhya* yoga develop an elaborate system of correspondence between the primal elements, the sense organs and perceptions – sound, colour, shape and form. Each element then begins to connote a specific colour and a geometrical shape.

The earth, for example, is associated with the colour yellow, and the shape of the square. Its emblem becomes the *vajra* and its purpose steadfastness. Water is white, the shape half-moon the emblem lotus and its purpose freshness. Fire is associated with the colour red, and the quality bright. Its shape is the triangle, its emblem a flame and its purpose burning. Wind is denoted with black, has the shape of a circle, its emblem is a dot and purpose impelling. Space is transparent, its shape a lotus, its emblem the void and purpose voidness. The tantric schema also establishes relations between the elements and the phonemes. There is a sizable body of literature on the subject.

The concept of the five elements is of fundamental importance in the schools of medicine. The body is considered to be made up

of the five elements, a continuum is established between the elements and the human body. Classical *Ayurveda* adopts the *Sāṅkhya* view of the transformation of matter.

It is these conceptions which determine the grand and schema of the *Tantras* and the *Agamas*. The artistic traditions both at the level of theory and practice, subscribe to the view of the elements and their conjointedness, the relationship of the sense-organs perceptions and the elements, and the levels of the gross and subtle. All these are then related to the *gunas* and are associated with sound, colour and shape. An Indian temple in its architectural ground and elevation plan, its structural edifice as also its configurations of visual imagery in the form of the vegetative animal, human life and its symbolic geometrical forms and shapes is the most concrete example of the transformation of speculation, ideational thought through a language of art.

We have spoken at the very outset of the concern with the three qualities of *tapas, ṛta* and *satya*. The structures created by the conception of the *mahabhutas* and their relationship and correspondence with the *indriyas* is this quest for observing and participating the being both the experiencer (*bhokta*) and a seer (*drasta*) of the process of creating and recreating an order. The order (*ṛta*) is not static, it is always in flux creating states of equilibrium in space and time. The foundations of the system provides a flexibility which facilitates multilayering and polyvalence.

It will perhaps be clear from the brief and simplified narration of the categories of one, two, three and five, that each of these categories are explored at the material and the non-material, at the tangible and intangible, micro and macro levels. The cojointedness and connection between and amongst them is a basic concern. Further each group is related to another group and finally a variety of grids are evolved. An elaborate a system of correspondence is established between the entities of one group and another. The system also evolves codes which are mathematical, however they are often couched in a language of myth and narrative. It is obvious that in such a system there can be no dichotomy between the sciences and philosophy, between the purely speculative and the theoretical and the praxis. We could detail the categories of division into seven and nine, but perhaps this is now not necessary.

The methodology of discourse is to identify a concept like a point, an almost dimensionless point and then allow it to expand in a series

of concentric circles. It is this methodology, which gives a central importance to the concept of the bindu the point. Indeed the concept of *bindu* is related to the concept of *bija* (seed) which denotes organicity, sprouting, growing from a single source, the emergence of a tree with trunk, branches, leaves, flowering, and fruiting and again seeding. The seed which grows into a plant from a single source. The metaphor of the seed, growth fruitation and seeding is used in the tradition in all domains. It denotes organic growth of concepts as also allows for renewal. The bindu is equally important. It has the nature of a drop of oil falling on an absorbent surface. In its very nature it expands. Different shapes and forms are created, but the most common is a kind of circle.

The worldview and the knowledge system is also like the *bindu* and to expansion with fluidity. Liquidity and flow are the characteristics of its natural behaviour (*svabhāva*). The systems we have spoken about also have this quality of a central point, a single source of undifferentiation and a gradual expansion. The *chakra* (wheel) with a centre, telly, spokes and circumference as a key metaphor of discourse was only logical. This methodology of discourse is in deep contrast to another system which develops fixed categories and of frozen structures, and places them in a linear progressive order.

Our unconventional and simplified account of the knowledge system and its structures would not be complete without at least a passing reference to the categories which move in a series of even numbers, and not odd numbers outlines above.

We are aware of the division of the directions into four, six and eight of the goals of life into the four purusarathas and the life-cycle into the four asramas. Each of these suggests generally a horizontal dimension and is considered to be the sphere of human behaviour conduct. The directions - disa intersect with the categories of loka we have spoken about.

Without elaborating further, it will perhaps be clear that the system has the potential of great complexity along with fluidity and flexibility. Acceptance of the movement from undifferentiation to differentiation is basic. So also is the consciousness of interdependence and reversibility and concords. The evolution of a series of correspondence of levels, domains of the material and non-material is the norm and not the exception. This is the *satya* to be pursued through a multiplicity of paths, the *ekam sad viprah bahudha vadanti* of the *Ṛg Veda*. It is also the movement of the journey of the no-

form no category (*arupa*) to a form and category and many categories *rupam, rupam pratirupa* to the beyond categories to an experiential consciousness of beyond form. In short this is an open and not a closed system.

In conclusion we have to return to the conversation between Tagore and Einstein and the comment of Ilya Prigogine. Tagore asserted that there was no universe other than the human. Ilya Prigogine has said "the basic interrogations are common to modern science and to Indian civilization", and Roger Penrose while responding to some questions on mathematical logic particularly Godel's theorem and models of the universe had responded thus to Ranjit Nair.

> The basic argument is that mathematical understanding, I would like to use the word 'understanding', as the key word is my own considerations, is beyond the scope of any computation. Take any computer activity; no matter how cleverly it can perform elaborate functions, there is no understanding present in that activity. Understanding requires awareness. It is one feature of our awareness, and that is the only one I can get a handle on."

And what is this under-standing if not the capacity of the human mind and consciousness.

In regard to the models of the universe as closed or open his preference was for the open one. He was of the view that the open model from a geometrical point of view is the most interesting. Indeed non-Euclidean geometry worked, was more consistent. He went on to add, "It now seems that these geometries are actually very beautiful, more than Euclidean geometry. As far as the aesthetic qualities are concerned, I would say the open universe is the most mathematically appealing". And what would be the precise definition of these aesthetic qualities ?

The world of concepts and the knowedge systems outlined above emerge from understanding the universe as an open one, with the complexity of going beyond computation, and following a non-Euclidean geometry which is as aesthetically beautiful as challenging. If perceptions of mind and consciousness allow the system to be both expansive and also intense, these are different pathways of the inexplicable mind, which are manifested in the sciences, philosophy and the arts.

Exploring Mind and Consciousness

Theoretical Approaches

The Mind-Body Problem in Three Indian Philosophies
Śaṅkara's Advaita Vedānta, Gaṅgeśa's Navya Nyāya, and Aurobindo's Theistic Monism

STEPHEN H. PHILLIPS

What is the mind-body problem? What's the complaint? Knowing the difficulty, we can better assess advice to be garnered from Indian philosophies. A theme of this paper is that there is no single mind-body problem but a family of related difficulties concerning the relationships between consciousness and matter. Like the proverbial lump in the rug, philosophers move the problem around, and none, I think, avoids it altogether.

Probably the best starting point is the famous "explanatory gap" from a scientific point of view concerning what is called phenomenal consciousness. Thomas Nagel in a famous paper entitled "What's It Like To Be a Bat" (1974) distinguished (a) our sense of our own experiences from the inside from (b) understanding even the most perfect description or theory of these that did not presuppose personal acquaintance with that which is being described. According to Nagel, not even the best theory could make us understand what's it like to be a bat from a bat's perspective. Although there are research programs that target cognition and thought and other mental abilities, no one seems to have a clue about how our experiences as they are to us from the inside might be explained.

Briefly let me rehearse the primary positions on the impasse.

(1) Eliminativism denies that there is any such thing as consciousness, the term "consciousness" and others such as "desire" and "will" belonging to a folk psychology that will be replaced eventually by proper science, the terms destined to disappear from scientific discourse like "flogiston" and "ether."

(2) Reductionism (comprising physicalism, functionalism, and similar views) acknowledges the reality of consciousness, but

insists that the explanatory gap is not unusual, that the biology of the future will understand consciousness just as fully as photosynthesis is understood now.

(3) Naturalistic non-reductionism finds no reason to regard consciousness as a non-natural phenomenon but concludes that human concepts are for various reasons inadequate to the explanatory task. And, finally,

(4) dualism, transcendentalism, and all views proposing that consciousness is non-natural or at least non-physical. All three of the Indian philosophies that I propose to survey fall into this last category, but their strengths are very different.

First, Advaita Vedānta, where consciousness is the primary reality, at least consciousness of a certain sort. Advaita, like all Vedānta, is a spiritual philosophy, supporting yoga and other consciousness disciplines so that, we might say, the primary reality might be better manifest. Now recent literature on the mind/body problem has identified several distinct types of consciousness. Phenomenal consciousness, access-consciousness, monitoring-consciousness, and self-consciousness, for example, have been shown to be different through assiduous intellectual labor in what is now a major subfield of contemporary philosophy. The type of consciousness that Advaita extols, in contrast, has not been given much attention—not, that is, by philosophers ignorant of Indian traditions who, unfortunately, are numerous in the United States. Now this type of consciousness might be called self-consciousness, as it is indeed in Sanskrit by its champions. But in recent philosophical literature, self-consciousness is usually talked about as the possession of a concept of self or as the ability to use such a concept in thinking about oneself, whereas the self-conscious-ness upon which Advaita focuses—an immediate and intrinsic, "non-dual" awareness of awareness—is said to be non-conceptual and independent of all thought even that about it. It is also supposed to be independent of material determinations and states.

What would an Advaitin have to say about the quandary known as the mind-body problem? In brief, the Advaitin would see the gap between science and this non-dual consciousness as uncloseable because self-consciousness according to Advaita is self-illumining consciousness and, if not unconnected to material states, is connected in a way that cannot be determined in thought. Self-illumining consciousness is inaccessible to representation and all

third-person point of view. The central plank of Advaita philosophy is the transcendence of self-illumining consciousness. And the self-illumining nature of this consciousness secures both a compatibility with science and an insularity from scientific explanation, so an Advaitin would argue, as I will explain.

First, some terminological clarification. Even if this were an audience of classical Indian philosophers, use of the term "self-consciousness" would be contentious, for there are many different views of self-consciousness in Indian philosophy. Thus what the Advaitins are about is better termed "self-illumining consciousness," an expression found in the earliest strand of the Vedānta tradition, the *Bṛhadāraṇyaka Upaniṣad* (c. 800 BCE). The usage occurs in a passage about transformations of consciousness in dream and mystic trance. A person is said first to dream by her "own light" and then to become "self-illuminated," *svayaṃ jyotiḥ* (4.3.9). Light is chosen as an analogue apparently because light illumines itself. A lamp illumines objects other than itself but does not require another lamp to be itself seen. Furthermore, in the *locus classicus* for Advaita views, in the works of Śaṅkara (c. 700 CE), namely, we find the insistence that this consciousness is "non-dual," *a-dvaita*, that is to say, that it knows itself by being itself, knows itself non-reflectively in a non-intellectual and indeed non-observational manner. Thus "self-illumining consciousness" seems the appropriate expression, capturing what seems to be kind of a phenomenal self-content.

In other words, what the Advaitins appear to have in mind is a phenomenal consciousness whose content is itself. This is nevertheless supposed to be a *state consciousness*, not a *consciousness-of*, not a transitive consciousness but a, so to say, intransitive one. Alternatively, we could say that this is a *consciousness-of* in a sense; it is a consciousness of itself. So here the *consciousness-of* relationship would have to be understood not as the asymmetrical relation it is normally taken to be but rather as something like identity. Note that in the Advaita understanding of this as self-consciousness, "self" is not taken to refer to the body or even the person but rather only the consciousness that is self-aware.

Discussions of types of consciousness normally proceed by presenting examples that are analyzed as exhibiting the one type in contradistinction with others with which it might be confused. But here, by the admission of distinguished members of Śaṅkara's very own school, the best that can be done is a dubious so-called

"indicatory" or ostensive definition. Directions are given where self-illumining consciousness may be found (e.g., the injunction, "Meditate"), which are said to be like a phenomenal definition of "red" that describes conditions under which one would normally experience the color. Now I say this indicatory definition is dubious because in Advaita's classical adversary school of Nyāya or "Logic" there is, I think, cogent criticism of the move, as I will explain later. Nevertheless, Advaitins claim that this consciousness is available to anyone capable of understanding the directions ("Meditate") such that consciousness attends to itself. But there is no third-person access to this consciousness, so it is said. It is denied that it shows itself in action. The gap between this consciousness and everything else makes discourse that seems to be about it problematic—perhaps the tell-tale weakness of the Advaita view, as I shall argue in accordance with the Nyāya criticism. For the moment, however, let us ignore the difficulties about language.

For, Advaitins do contrast self-illumining consciousness with other types of consciousness or cognition, *jñāna*, presupposing, it seems, an audience of compatriots who know directly, or who could know, what this consciousness is. That there is an epistemic specialness to self-illumining consciousness is said to be proved by a sublatability argument. Now this argument is not put forth to show that the capacity of other types of consciousness to guide action should be questioned; Advaitins are mislabelled illusionists and skeptics. They do not deny the epistemic value of perception, for example, but use epistemic terms to distinguish self-illumining consciousness. The *contrasting* argument, then, goes as follows.

Everything dualistically experienced is at risk of being shown to have been mispresented through experiential sublation. Thus all cognitions could be sublated and shown to be non-veridical—except self-illumining consciousness. For, since among cognitions only self-illumining consciousness is not an appearance of one thing in or qualifying another, only in its case is a precondition for sublation not met. That is to say, a perceptual illusion can be sublated because a perception presents an object as qualified by or bearing a property. (In Sanskrit it is said that the property is cognized as *located* in the property-bearer as well as, in an alternative expression, that the object is cognized as *qualified* by the property.) And a property presented may not qualify the object in fact, as in the case of silverhood presented as qualifying what is in fact a piece of mother-of-pearl.

Perceptual cognition is qualificative; something a is cognized as F. Sublation is an ensuing experience that shows the a not to be F. But self-illumining consciousness is not in this way "qualificative," to render the Sanskrit expression, *vaiśiṣṭya*. Its non-dualistic mode of presentation precludes sublation, in sum. It cannot possibly be non-veridical—unlike all perception and indeed all thinking (remembering, inferring, understanding what someone has said, et cetera), all normal cognition which is invariably consciousness-of of the transitive type. Thus self-illumining consciousness is self-authenticating, at least so Advaitins say. Interestingly, Advaitins also say that there can be no real point to any question about authentication. To itself, self-illumining consciousness stands self-revealed.

It is not my intention to pursue today the sublatability argument. I mention it mainly in the contrasting spirit mentioned above. Still, we do need to know that Advaitins also extol the value of self-illumining consciousness along such lines. All contrasts drawn serve the agenda of encouraging people to attend to this. Indeed, the Advaita school is defined by its commitment to the reality and value of self-illumining consciousness. What I want to show now is how this connects with a theoretical minimalism and Advaita's attitude towards science.

Advaitins over their long history—from Śaṅkara to modern advocates such as Sarvepalli Radhakrishnan—have been able to align themselves with distinct theories about the operation of the sense-organs and the generation of cognitions through physical processes because of a great merit of the Advaita view that Advaitins themselves point out. This is that the commitment to self-illumining consciousness is compatible, at least so goes the thesis, with all science and externalist theory, except, it seems one has to add, that which would propose to explain self-illumining consciousness itself. More about the exception shortly. First it is important to see that there is no call from the Advaita side to explain the world in relation to self-illumining consciousness. For, self-illumining consciousness is self-contained, whereas explanation presupposes a relationship between explanandum and explanans. An explanation would employ terms that are learned through ostensive training by teachers of pupils and the customs of everyday life. Self-illumining consciousness cannot serve even as an explanatory first principle since it is to such terms and teaching unavailable. As we have already learned, the

radical internalism of Advaita about self-illumining consciousness is said to have the consequence that speech can only indicate it, not refer to it directly. We'll come back to this claim about language. The main point now is that self-illumining consciousness is non-relational whereas an explanation would purport to find a tie between explanandum and explanans. Self-illumining consciousness does not explain anything.

Nor can it be explained, for the same reason. That there is thus mystery in consciousness revealing diversity, and in the transition from the one to the other, is readily admitted by Advaitins. That is to say, why there are both self-illumining consciousness and the worldly display is part of what is inexplicable, *a-nirvacanīya*. This thesis flows from the non-relationality of self-illumining consciousness and the nature of explanation as supposing a relation between explanandum and explanans.

The list of great Advaita philosophers includes several who are expert at dialectics, expert at finding counterexamples and disputing philosophic theories. Often one can find no motive behind the Advaita refutations other than sheer sport. And, to come to the point, Advaitins typically embrace the science of their day (although there are exceptions). Nevertheless, there are, from the Advaita point of view, theories—let us call them metaphysical theories—that overstep the bounds of science and purport to explain self-illumining consciousness itself. Invariably, they try to do so in relational terms, try, that is, to integrate an understanding of "self-consciousness" into a holistic theory. Such theorists, say the Advaitins, necessarily make a mistake, because they treat self-illumining consciousness as related to other things.

In this context, it is important to evoke an epistemic perspective, because epistemology provides the canons of winning and losing a debate or dispute and we are now engaged in a dispute, so say champions of the Advaita cause. Self-illumining consciousness is self-authenticating and, unlike other conscious states and material phenomena, has an exclusive access to itself. Thus only it has the right to pronounce on itself, so to say. Of course, about itself it has nothing to say—it does not speak—but it knows itself directly. Those who would explain self-illumining consciousness are trespassers in this domain invariably misrepresenting an explanandum.

Personally, however, I find the non-relational thesis that underlies

this line of argument unsatisfactory. My second concern, the classical philosophy of Nyāya, convinces me that there is an impossible tension at the core of the Advaita stance. Just how is it that Advaitins know—such that they may out loud say—that self-illumining consciousness cannot be related to other things? The idea seems to be that this consciousness's being absorbed in itself translates into its being explanationally unavailable. An explanation would be like an unwanted disturbance violating self-illumining consciousness's self-absorbed trance. But how is it that Advaitins know so much about this "self-absorption" to say "self-absorption," "self-illumination," and the other things they do say including suggesting (but at the end backing away from) why this consciousness should be inexplicable? This brings us back to the problem of the language Advaitins use.

The insularity from other areas of theory is purchased by Advaita too cheaply. Advaitins draw a distinction between attributive use of words and the indicatory, as I mentioned earlier. There is supposed to be a difference between (1) describing Devadatta's house, saying what a house is—say, to a child—and (2) indicating Devadatta's house conversationally by saying that it is the one where some crows pointed to are hovering. The same distinction is now well-known in contemporary philosophy of language, the referential/attributive distinction—which, then, let me just for one moment digress, was known in India more than a thousand years earlier, being used, moreover, by philosophers again and again in various contexts and arguments. Indicators (*upalakṣaṇa*) and qualifiers (*viśeṣaṇa*) are to be distinguished, so goes a prominent theory, in relation to negations to be made known. More precisely, negations, such as "(It's a cow, therefore) *not a horse*," can be inferred from a qualificandum, a property-bearer (the cow), brought into focus by either an indicator (hovering flies) or a qualifier (cowhood) as a bit of express "predication content" of an experience. In the former case, the relation between the predication (hovering-flies) and the inferrendum (not-a-horse) is indirect, whereas it is direct in the case of the latter, where the true qualifier (cow-hood) is presented. In the case of the hovering crows pointed to to indicate Devadatta's house, that the house does not have, for example, the shape of a lotus pond is an inferrendum specified by a *qualifier* presented as predication content (the shape of Devadatta's house), whereas it is not an inferrendum specified by an *indicator* presented (the hovering

crows). Thus the distinction is an experiential version of that between linguistic attribution and mere reference—as when we say that Jones is Smith's murderer (*attributing* murderousness to Jones) in contrast to saying that Smith's murderer (i.e., Jones, who is on trial) is innocent (using "Smith's murderer" merely *referentially*, merely to indicate Jones, like the hovering crows mentioned to indicate Devadatta's house). So understood, the distinction cannot be used by Advaitins in the way they try, so argue the Nyāya realists and referentialists and, I must say, I agree.

The distinction is that between true and pseudo-qualifiers in connection with use of words to pick out something by way of a property it really has as opposed to words used to direct a hearer's attention to something by means of a thoroughly contingent and accidental relation it has to the words' literal and direct referendum. But, so Nyāya philosophers argue, the distinction provides no help to the Advaitin in his trying to direct our attention to self-illumining consciousness, because even mere pointing implies a relation, a line connecting, however ephemeral, the sign and the signified. In the reverse direction, in what appear to be attributive usages, Advaitins consider themselves to know some true properties of self-illumining consciousness. Why else would they say that it is self-illumining, or non-relational, and then reason according to these attributions? We Naiyāyikas do find self-awareness, which is an awareness that takes another awareness as its object or indication. We do not find, however, any self-illumining consciousness, nor have we been told coherently where to find it.

Now there is a long and complex debate between Nyāya and Advaita over the concept of self-awareness and the question of how consciousness is itself known, to which Buddhists add arguments on the Advaita side along with Mīmāṃsakas of at least three different camps and to which, on the opposite side, on the side of the Naiyāyika externalists, are joined various voices in a diverse theistic camp. But over all this today we shall have to skip. Let us turn directly to the positive theories of Nyāya, my second topic, and to the question of the advice Naiyāyikas would give to those vexed by a mind-body problem. Advaita would purchase compatibility with science by maintaining that the "explanatory gap" cannot be closed. Yet by its own premises it fails to show us why it cannot be closed, why self-illumining consciousness cannot be explained. Is there a similar lump in the rug of Nyāya?

Nyāya is itself straightforwardly dualist in the sense of admitting the reality of distinctively mental properties and entities such as selves along with atoms and other material things. Indeed, Nyāya is ontologically pluralist, finding nine types of substance and several types of property as distinguishable ontological items. An awareness is a psychological property, for example, distinct from physical properties such as colors and shapes that belong to rocks and other things that are made of material atoms. An awareness rests or occurs in a self, which is another distinct type of entity, and only for an instant before giving way to another awareness, each indicating an intersubjective object or objects other than itself.

To move quickly to our question, Nyāya's peculiar take on the mind-body problem would be to deny that there is an explanatory gap. For this point and its elaboration, I draw principally from the classic text of Navya or "New" Nyāya, the *Tattvacintāmaṇi* of Gaṅgeśa (c. 1325), but Gaṅgeśa's followers, too, who comment on his views at great length, would also deny any gap. All these Nyāya philosophers formulate causal principles on the basis of positive and negative correlations without a bias about the sorts of things that can be correlated. For example, sensory connection with an object a that possesses a property F is found to be a cause—in the sense of a necessary condition—of a veridical and reliable perceptual awareness with Fa as its object. An induction is made relating mental entities, e.g., perceptions of such a type, and physical entities, the F-possessing objects that come to be in sensory connection. Correlations are made, similarly, between efforts with propositional content and certain physical, i.e., bodily, acts.

Thus, Nyāya has an ingenious strategy for the mind-body quandary, following the motto, "Look for correlations, but don't overinterpret. Be economical." Its a strategy that deserves close attention, but I, for one, do not think that all versions of the mind-body difficulty would thereby be side-stepped. For I think the problem shows up in Nyāya despite its distinctive ignorance of it and the plausibility of many of the mental/physical correlations it propounds.

First, a few further remarks of an introductory character. Nyāya's thrust, it is often said, is epistemological. This is right: the school is principally concerned with, we may say, warranted belief. But Naiyāyikas try to attune their epistemological principles to a picture of a world of things external to cognition and bound together by

causal laws. An awareness is a psychological property or event that is the effect of perceptual, inferential, and other physical/mental processes. And an awareness is itself a cause of the formation of memory-dispositions as well as—it may be, depending on other factors—of a particular line of action. As a psychological property, an awareness is distinct from the properties of a person's body and of the objects grasped by the external organs of sense. An awareness rests or occurs in a self, which is not a material entity. An awareness lasts only a brief instant, being replaced by another partly by means of the nature of a current awareness but also according to several further factors. For example, a new sensory connection such as of the organ of hearing with a loud sound can force its way into consciousness, overriding other factors such as a desire to focus on something else, a taste, for example. So, there are, according to Nyāya, causal continuities and processes involving entities that are physical and mental—in the one direction through the operation of the sense organs, sight, hearing, and so on, and in the other in guiding action, voluntary action, to be sure, including the speech act of describing something perceived, i.e., verbalizing an awareness's content or, technically, objecthood (*viṣayatā*). There is also a third type of cognitive causal relationship, a kind of mind/mind causation exhibited in the causality obtaining between two successive awarenesses.

A problem in the Nyāya picture that I wish to reveal, following the arguments of the modern Vedāntin Sri Aurobindo, is the presumed perceptual availability of properties previously perceived through retrieval by means of memory. There is here a kind of physical/psychological interaction that is typical of Nyāya in its formulations of cognitive laws. But Nyāya philosophers abandon their avowed direct realism in admitting an "extraordinary" connection in the case of a property perceptually presented through the operation of memory. The problem is that if properties can be supplied mentally and there is no phenomenological difference between those mentally supplied and those supplied physically, why not, in the spirit of Nyāya's own preference for economy, jettison the distinction? This could be either in the direction of the idealism of the Buddhists or in the very different type of realism promoted by Aurobindo, to expand the context of the classical controversy.

To repeat, Naiyāyikas tell us they are direct realists; a perceptual awareness whose objecthood is "F*a*" is veridical just in case the object

a grasped by the sense organ is F. Normally *a*'s being F is responsible, in part, for generating the perception whose object is F*a* though there are, as Naiyāyikas are aware, so called Gettier-style cases where the object that a pseudo-perceptual awareness is about is indeed an F but has not in fact generated the awareness verbalized as "*a* is F." These cases, however, will not be our concern here, but rather Nyāya's explanation of non-veridical perception. The story goes as follows in the case of an illusory cognition of silver when the object in front is in fact mother-of-pearl.

A perceiver has had previous veridical experience of silver that has formed a memory-impression, *saṃskāra*, that when revived for whatever reason (such a reason is called an awakener, *udbodhaka*) triggers a remembering that has silver as its object. Now when this perceiver sees a piece of mother-of-pearl as silver what has happened is that the memory-impression of silver has similarly been revived but instead of triggering a remembering has fused a previous perceptual content or objecthood into a current perceptual presentation of an object in front. The perceiver would verbalize the current perception as, "This is silver," and he would be right about the "this," right about there being an object in front that he is perceiving and referring to as "this." What is right about the perception concerns its qualificandum portion, as Naiyāyikas would put it; the object of the cognition qua property-bearer is successfully hit. But the qualifier or predication content portion presented (*prakāra* in Sanskrit, the "way" an object is perceived or what it is perceived as), here the being-silver, or silverhood, is not true of the object in front, and this "content" is furnished, says the Naiyāyika, not by the normal perceptual process but by memory. Still, the predication content from the perspective of the perceiver is thoroughly experiential. He reaches out for the mother-of-pearl that he regards as silver because he is perceiving it as silver. Being-silver is perceptually presented. The causal analysis of the illusion has it that the memory-impression, as opposed to the object's being silver in fact, plays a causal role. Note, finally, that this is to consider it all from a third-person or externalist standpoint.

Naiyāyikas call the sensory connection characteristic of illusion "extraordinary," *a-laukika*, and thereby signal their own version of a mind/body problem. The ordinary sensory connection provides Nyāya its chief link between body and mind. When we see a blue lotus, normally it is the lotus's being blue, its blueness, that is causally

responsible for the "blue" part of our perception. The causal process works crucially through a sensory connection with the qualificandum on which the qualifier rides piggyback, the qualifier, that is, that appears as predication content. In the case of a perceptual illusion "F*a*," there is still required a fact of Fs and indeed of our subject's having encountered an F previously; otherwise, he would have no memory-impression of Fs as required by the causal account. So there is in a sense a sensory connection with Fs in the case of the illusion "F*a*," though considering the F part, the predication content, such a connection works through memory and is thus "extraordinary." Illusion, claims the phenomenalist, whether Western or Buddhist, shows us sense data, shows us that the perceptual object or content belongs to the perception itself and not some worldly thing. Naiyāyikas refuse to countenance sense data or qualia; perceptions present the world which is intersubjective, not properties internal to perceptions themselves. Indeed, even perceptions and all awarenesses are real things, related to all other real things. They are properties of a self, and may be grasped in apperception which is a distinct cognition having a target cognition as object or content, what it is about. But by admitting that the sensory connection with silver is extraordinary in the case of a piece of mother-of-pearl misperceived that way, the Naiyāyika beholds the mental/physical gap that is part of what is called the mind-body problem, specifically that part of it that is the mystery of how physical objects somehow intended by sense data can be responsible for them.

To be fair, it should be pointed out that illusion is not the only case that the Naiyāyika identifies as involving an extraordinary sensory connection. The role given memory in the account of illusion is not entirely *ad hoc*; there are other, non-illusory instances of memory retrieval—or revival, perhaps we should say—of a perceptual content. These are illuminating. For example, so long as the fragrance of sandalwood has been previously experienced, even the visual organ can provoke an awareness of it, through this same type of mentalized sensory connection called by Naiyāyikas extraordinary. That is, a visual perception of a piece of sandalwood at several meters distance which would be verbalizable as, "A piece of *fragrant* sandalwood," is cited as a non-illusory example of the phenomenon. This would involve an "extraordinary" sensory connection in that the fragrance is said to be visually presented—to speak of the most immediate role of a sense organ—although, of course, the experience of the fragrance is of a smell. A memory created by a previous smelling experience gets fused "extraordinarily" and becomes perceptual content, although there is no current contact of the smelling organ with the sandalwood. Apparently this is borne out phenomenologically: people sometimes say they smell the fragrance of a piece of sandalwood that, in a physical sense, they only see. Thus, as in error, a memory-impression would play a crucial causal role.

The theory of extraordinary sensory connection has wider application still. It is invoked to explain recognitions, e.g., "This is that Devadatta I saw yesterday." On the Nyāya account, a recognition is a perception informed by a memory-impression as an auxiliary cause. In other words, the problem of how an object's thatness—its being that thing that was experienced yesterday—can be perceptually presented currently is solved through this same theory of extraordinary sensory connection, *a-laukika-saṃnikarṣa*. A current sensory encounter with Devadatta, who has been previously perceived by the person S now recognizing him, sparks S's memory to project into the perception the "thatness," and S says, verbalizing his perception, "This is that Devadatta." It is to be stressed that, for the Naiyāyika—and also for other classical realists—Devadatta is thought of as qualified by the thatness in fact. A recognition depends on memory for the "thatness" cognitively, but the thing recognized is the *that* in fact. The presentation of the *that* qualifier has a memory-impression as an auxiliary cause and is thus extraordinary, not the fact that Devadatta has thatness. So just as in the case of memory's reviving a smell previously experienced and making it a part of a current perception of sandalwood, we have here an example of a veridical perception where memory plays a causal role parallel to its role in illusion.

There is a lot one could say in criticism and also in defense of this theory, and a lot indeed has been said in the numerous texts of classical Indian philosophy. But let me draw out only the morale that I indicated at the beginning, namely, that the Naiyāyika faces his own version of a mind-body problem. Ordinarily, perception, according to him, is generated by a sensory connection with an object that is as it is perceived with no role for memory. There are cases, however, where this view is clearly implausible—no, worse, where it is clearly wrong. There is no sensory connection responsible for a rope appearing as a snake, no current connection with something's

being a snake. Similarly, there is no current connection with Devadatta's being the person who was encountered the day before, nor with the fragrance of a piece of sandalwood that is seen at a distance. Yet the sensory connection is a main bridge between the physical and the mental, according to Nyāya. There is also effort and action, where the bridge is the *manas*, the "internal organ"—which is a story that must be left for another occasion but a story that, I would contend, shows similarly a mind-body problem. My point is about perception, that the Naiyāyika in his own way recognizes there the mystery of physical/mental connection by calling the cases extraordinary where memory, by his account, plays a causal role. For, by his own admission, the mind—i.e., memory—is just as capable of generating perceptual content as something physical. There is nothing about a perception itself that insures a connection to the physical world as opposed to something furnished by one's own mind, or memory, which, by the way, is another hypothetical posit, there being no direct experience of it but only of its effects in apperceptions of rememberings and other cases. Thus Naiyāyikas have to give mysterious powers to memory to save their dualistic realism and their laws that run, in this instance, from body to mind.

A third Indian philosophy combines the realism of Nyāya about the objects of perception (including matter and mind) with the spiritualism and emphasis on consciousness that is characteristic of Advaita and indeed all Vedānta. The theistic monism of Aurobindo Ghose (1872-1950) embraces both Advaita's commitment to the transcendence of consciousness and the realism of Nyāya about the physical world, and would close the explanatory gap by reversing the terms of the consciousness/body relationship. Consciousness becomes the explanans. That consciousness self-determines is Aurobindo's byword. Consciousness by nature can make itself material. The universe is the body of God—says this early twentieth-century thinker who thus associates him-self with a family of philosophies of theistic Vedānta, the closest of which is probably Rāmānuja's Viśiṣṭādvaita Vedānta.

As transition to Aurobindo, let us consider another dimension of Nyāya's causal realism, its theology. Later Naiyāyika authors in particular are much occupied by a causal argument to God. God is inferred as the agent who has brought about things like the earth. Nevertheless, just as Nyāya does not speculate on what makes it

possible for mental and physical events or properties to stand as causal factors one for the other, so Nyāya refuses to speculate on the connection between God and the finite things that directly or indirectly God brings about. To be sure, along with the inference to God, the later Naiyāyikas attribute omniscience (*sarvajñatva*) to God as a property God must have to bring about the effects the Īśvara is posited to explain. But it seems to be school policy not to elaborate, not to speculate further on God's nature, nor—in the words of an objection often discussed in the classical literature—to say how it is possible that God, who is generally assumed not to have a material body, can create earth and the like which are material things. Furthermore, some things God does not create, such as individual selves, which are eternal and uncreated. Because of these and other features of the view, some contemporary interpreters have judged Nyāya's theistic dimension as on the whole peripheral to its central tenets. Aurobindo, in contrast, provides rich theistic theory. Probably it would be hard to imagine any metaphysics richer than his. He speculates on the deep connection between spirit and matter, of which God's creative activity is an instance.

Now Aurobindo does indeed use the Advaita concept of self-illumining consciousness as a starting point for his attempt to explain matter in spiritual terms. He says consciousness by nature "knows itself by being itself," and "knowledge by identity" is not only part of his view of the essential nature of the Absolute Brahman (who is God) but it also lies at the center of Aurobindo's mystically psychological picture of human consciousness. However, Brahman' *sva-rūpa*, Brahman's "essential nature," includes more than self-knowledge in the Advaita sense of the term. It includes power, power to self-determine. What's extra in Aurobindo's metaphysics is a view of Brahman as *cit* (or *cit-śakti*, "consciousness-force") on a yogic model. Let us focus on this and come back later to the notion of knowledge by identity and the inclusion of the Advaita stance.

"One-pointed-ness of concentration" (*ekāgratā*) is crucial to yogic accomplishment, according to authorities from Patañjali's *Yogasūtra* to Aurobindo himself. Aurobindo views Brahman as "exclusively concentrating" in the process of creation, which he calls involution. Creation is self-manifestation that proceeds through self-contraction. The Absolute is continuous in its being with the universe that it manifests—the continuity accounting for the deep possibility of our own self-knowledge, our own knowledge by identity, as will be

explained—but in manifesting Brahman's essence becomes implicit and submerged. That is to say, while Brahman is, as with all classical Vedānta, essentially infinite *saccidānanda*, "Existence-Consciousness=Force-Bliss," Brahman wills on itself a process of contraction, a process of "involution," to use Aurobindo's word, whereby God puts aspects or characteristics of itself behind a veil, through involution, to bring our world about. This involution is modeled psychologically on yogic "exclusive concentration," *ekāgratā*.

That is, "exclusive concentration," which is in yogic traditions a power of consciousness responsible for *siddhi*-s as well as for a person's achieving *samādhi* or "mystic trance," is for Aurobindo the analogical springboard to his understanding of Brahman as *cit*, which he renders as "consciousness-force" (*cit-śakti*). In Sanskrit, the word "*ekāgratā*" is an abstract noun formed from *eka*, "one," and *agra*, "point" or "tip," thus "one-pointedness." In Patañjali's *Yogasūtra* (c. 300 CE ?), it is identified as the mental or volitional quality of concentration of attention, as in meditation on a single object, without the mind wandering, without admitting distraction (e.g., *YS* 3.11 and 3.12). And, as mentioned, yogic authorities view *ekāgratā* as key to mystic accomplishment, to the psychological breakthrough called *samādhi*, where consciousness knows itself as it is essentially.

Consonantly, a theme of Aurobindo's yogic teaching is that yoga is power. He argues that our ability to concentrate one-pointedly helps us accomplish whatever we want to do, whether it be something physical like carpentry or mental like writing a book. Further, he provides the psychological gloss on the everyday phenomenon of exclusive concentration that it involves a putting into the background other concerns of ours or other aspects of our personalities, without our ceasing to value those other things or our ceasing to be the fuller person. The batter who thinks about his wife misses the pitch, but the one who hits it, not letting his concentration stray, should not *ipso facto* be accused of not loving the wife. He simply cannot think about her and bat successfully at the same time. The concentrated action requires a willed ignorance, Aurobindo would say. Thus according to this understanding even everyday concentration involves *a-vidyā*, a "not-knowing" with respect to Y and Z, or a not-being-aware of Y and Z, that permits a full awareness, and power, concerning X. Thus it is that Aurobindo would connect the power of consciousness with *a-vidyā*, a notion that reverberates in his metaphysics, as I will now elaborate, as it does also, albeit

interpreted differently, in previous Vedānta.

As mentioned, Aurobindo sees Brahman as "exclusively concentrating" in the process and maintenance of creation, in "involution." In essence, perfect Being, Consciousness-Force and Bliss or Value (so Aurobindo renders the traditional characterization, *sac-cid-ānanda*), Brahman involutes aspects of itself—that is to say, contracts or becomes purposely ignorant, assuming an *a-vidyā*—so that certain finite possibilities can emerge. There is thus no mind-body problem because self-manifesting Brahman remains a unity; Aurobindo's view is as much a monism as any materialist reductionism. Brahman's own exclusive concentration is the creative process linking spirit with the emergence of matter, Brahman's way of being each of us and all finite things. Involution has an outer limit in the "inconscient" energies of matter, where Brahman's native awareness and bliss are put almost entirely behind a veil. This is the cosmic *a-vidyā*. But since Brahman is essentially consciousness, the inconscience of matter can be only apparent, and an evolution of consciousness is natural wherever there is the chance. Matter cannot everywhere remain inconscient because it is nothing but Brahman. Thus there is an evolutionary nisus and a telos to cosmic biological process which is founded in the *cit* of Brahman.

To continue with the metaphysical picture, let me stress that, according to the great yogin, nothing essential to Brahman, though it be put behind the veil by Brahman's power of exclusive concentration, can remain behind the veil forever. This is perhaps Aurobindo's central argument, repeatedly formulated with different emphases. One way to interpret it is to say that conceivable universes that are incompatible with Brahman's essential nature are strictly impossible. A universe of absolutely insentient matter Aurobindo judges incompatible with Brahman. Thus living, conscious material beings are destined to evolve somewhere (indeed everywhere there is the chance). To move very fast to the future that Aurobindo envisages for our planet, the philosopher claims that many of us or our progeny are destined to know themselves as Brahman. God works within limits, and could not, for example, make 2 + 2 = 5. God could not create an entirely insentient world since God is constrained by the metaphysical law *ex nihilo nihil fit* ("nothing from nothing") to create out of God's own nature of Consciousness and Bliss ("How can being arise from non-being?" *katham a-sato saj jāyeta*, *Chāndogya Upaniṣad* 6.2.2). Thus this world is destined, he says, to evolve sentient

material beings and eventually a divine life conceived as a society where many have a rather direct experience of themselves as Brahman.

In sum, creation, according to Aurobindo, is self-manifestation that proceeds through an exclusive concentration, an involution, a kind of self-contraction. Brahman is in essence infinite *sac-cid-ānanda*, "Existence-Consciousness=Force-Bliss." It requires an involution—that is, God's putting aspects or characteristics of God's essential nature behind a veil through a willed ignorance, *a-vidyā*, psychologically modeled on yogic *ekāgratā*—to bring our world about and to sustain it. This process reaches its outer limit in the apparently inconscient energies of matter. But the consciousness involved in matter has to evolve out of matter eventually into a "divine life." According to Aurobindo, the inconscience of matter is no more intrinsic than the ignorance involved in scenes of exclusive concentration in ordinary human life. Only the immeasurably great Brahman, the supremely conscious, could so exclusively concentrate as to have become the apparently inconscient material universe, but not even Brahman can hide itself from itself for-ever. Matter reflects Brahman's power, but through its spiritualization in persons such as ourselves other aspects of Brahman become manifest.

Now to the mind-body problem. According to Aurobindo, Brahman's being matter explains how it is possible that we who are material beings can know ourselves immediately, self-illuminingly, as taught in Advaita. Matter's being Brahman also explains how there can be the sensory connections that Nyāya seems correct to identify but struggles with, as we have seen, as well as, to quote Aurobindo, how "genes and chromosomes [can be] the cause of hereditary transmissions, not only of physical but of psychological properties" (1973, p. 299), a possibility that is not explained by science. Similarly, it ceases to be a mystery, again to quote Aurobindo, "how a fixed formula for the combination of oxygen and hydrogen comes to determine the appearance of water which is evidently something more than a combination of gases, a new creation, a new form of substance, a material manifestation of a quite new character" (1973, p. 298). Since matter is consciousness involved, material forms can be the vehicles of mental determinations and all the emergent phenomena that Aurobindo intends to suggest by means of these and other examples.

Interestingly, we may go on to note, Aurobindo's philosophy of

Brahman as *cit-śakti* allows him to put the Advaita notion of self-illumining consciousness to new use. Knowledge by identity makes possible conscious control. By knowing myself as my hand I can move it, et cetera. A primary teaching in the *Yogasūtra* is that through *saṃyama*, "control through conscious identification," instanced in the control that we have over our own bodily limbs, we can expand the sphere of things subject to volition, as in consciously regulating the breath (*prāṇāyāma*, producing well-attested results for health) and in achieving, so says the *Yogasūtra*, such *siddhi*-s as reading minds and provoking friendship among all in the yogin's immediate environment. We might also mention here craftsmanship and athletic ability, which are developed through training, through conscious attention to what Aurobindo would see as an extension of our essential selves to handle expertly external tools as well as our own bodies and minds. ("Yoga is skill in works," says the *Gītā*, in a line of which he is fond, *yogaḥ karmāsu kauśalam*, 2.50.) Aurobindo's take on the mind-body problem would be to emphasize possibilities—which he would call yogic—to develop our consciousnesses through discipline (yoga) and to extend the sphere of the voluntary. In his teaching, the extension should occur in two directions, outward in mastery of matter and inward in mastery of desires and our less conscious parts along with integration with our higher selves. Thus we see the spiritual thrust of Aurobindo's metaphysics, the values that flow from it. Let us just quickly note how different these are from the rather nihilistic sides of the materialisms that are its monistic rivals.

The lump in Aurobindo's rug is of course evil. The explanatory gap is closed in his philosophy at the cost of making Brahman both responsible for evil and the bearer of it. For Brahman both creates and maintains the universe in its all-comprehensive being and knowledge. Now Aurobindo acknowledges the difficulty and devotes hundreds of pages of his master-work (1973) to accounting for it without abandoning his rich theology. Unfortunately, this effort on his part is itself much too rich for us to do more than sketch. My main point is simply that the difficulty is for Aurobindo severe. He reverses the explanatory relationship and thus encounters not the problem of explaining consciousness but of explaining how our life and world could be so bad. Given Brahman, our world of Brahman's manifestation should be a lot more glorious with not so many foul spots and so much suffering. How could a theistic Absolute, as the

Buddha argued, allow disease, old age, and death?

Evil, answers Aurobindo, is rooted in the insentience of matter and the limitations it imposes on life. These, then, are to be just the converse of valuable possibilities matter secures, namely, ourselves and the good things of life and the value of things future. In fact, the value of ourselves and this universe as we are now is not quite, in Aurobindo's conception, valuable enough. There has to be further evolution, development of individuals with finite bodies and minds who are nonetheless aware—mystically, yogically—of Brahman, who do not cease thereby to be individuals materially embodied but who no longer suffer from the fundamental spiritual ignorance (*a-vidyā*) of failing to know ourselves as manifestations of Brahman. As I suggested earlier, Aurobindo does not believe that a world like ours without further development in the direction of the grand telos he envisages would even be possible. That is to say, in Aurobindo's estimation our world as it is now, fixed in all its evil, would be incompatible with the reality of Brahman and thus could not be. Thus he takes evil as a sign of future evolution. In considerable measure it is indeed the value of future "divine life" that discounts, he reasons, the evil made possible and even necessary by the insentience of matter inasmuch as matter and involution make possible our evolutionary world. Biological evolution is for Aurobindo not merely a matter of life transformations but of progress, teleological change. Thus the fundamental nature of matter has to include—for metaphysical reasons—an "evolutionary nisus" that insures the emergence of individuals capable of mystical experience in which the supreme reality, Brahman, is revealed. This notion of an evolutionary nisus, or urge or material drive, is the bottom side of Aurobindo's fresh concept of Brahman.

Thus Aurobindo in facing his own version of a body-mind difficulty again relies on the conceptual tool of *a-vidyā*, "ignorance," as the necessary complement of the "exclusive concentration," *ekāgratā*, whereby Brahman self-manifests and creates our and any world of finite things. Ignorance is required for creation and maintenance of our universe, and thus Aurobindo would get a step up on the task of discounting evil. It only superficially seems that on the premise that the supremely real is Brahman, *saccidānanda*, things should be a lot better here. In this way, Aurobindo marshals, I would say, his best resources to the explanation of evil. It is not, however, my point today to try to evaluate this effort, though I will say that the approach would

have to be a bit strained, it seems, in certain applications. Again, my main point is just to point to the lump. The final question seems to be, then, where one prefers to find it.

BIBLIOGRAPHY

Aurobindo, Sri (Ghose). *The Life Divine.* Pondicherry: Sri Aurobindo Ashram Trust, 1973.

Brahmasūtrabhāṣya of Śaṅkara. With the *Bhāmatī* commentary by Vācaspati Misra. Ed. J. L. Shastri. Delhi: Motilal Banarsidass, 1980.

Chakrabarti, Kisor. *Classical Indian Philosophy of Mind.* Albany, New York: State University of New York Press, 1999.

Nagel, Thomas. "What Is It Like To Be a Bat," *Philosophical Review* 83, 1974, pp. 435–50.

Nyāyasūtra (Nyāyadarśanam). With four commentaries, the *Nyāya-sūtra-bhāṣya* of Vātsyāyana, the *Nyāya-sūtra-vārttika* of Uddyotakara, the *Nyāya-sūtra-vārttika-tātpāryaṭīkā* of Vācaspati Miśra, and the *Vṛtti* of Viśvanātha. Ed. A.M. Tarkatirtha, Taranatha Nyayatarkatirtha, and H.K. Tarkatirtha. Calcutta Sanskrit Series 18. 1936–44. Reprint, New Delhi, Munshiram Manoharlal, 1985.

Phillips, Stephen H. *Aurobindo's Philosophy of Brahman.* Leiden: E. J. Brill, 1986.

———. "The Conflict of Voluntarism and Dualism in the *Yogasūtra.*" *Journal of Indian Philosophy* 13, No. 4, Dec. 1985, pp. 399–414.

———. *Classical Indian Metaphysics.* Chicago: Open Court, 1995. (Indian edition) Delhi: Motilal Banarsidass, 1996.

———. "Two Problems about Perception and Mental Intermediaries in the Nyāya Dualism: Focus and Extraordinary Sensory Connections with Perceived Properties," *Journal of Indian Philosophy and Reli-gion* 5, Oct. 2000.

Radhakrishnan, Sarvepalli. *An Idealist View of Life,* 2nd ed. London: George Allen & Unwin, 1937.

Tattvacintāmaṇi of Gaṅgeśa, vol. 1, *pratyakṣa-khaṇḍa.* With the *Prakāśa* commentary by Rucidatta Miśra and a subcommentary by Rāmakṛṣṇādhvarin. Ed. N.S. Ramanuja Tatacharya. Kendriya Sanskrit Vidyapeetha Series 20. Tirupati, 1972.

Tattvacintāmaṇi of Gaṅgeśa, vol. 2, part 1, *anumāna-khaṇḍa.* With the *Prakāśa* commentary by Rucidatta Miśra and a subcommentary by Rāmakṛṣṇādhvarin. Ed. N.S. Ramanuja Tatacharya. Kendriya Sanskrit Vidyapeetha Series 33. Tirupati, 1982.

Tattvacintāmaṇi of Gaṅgeśa, vol. 2, part 2, *anumāna-khaṇḍa.* With the *Prakāśa* commentary by Rucidatta Miśra and a subcommentary by Rāmakṛṣṇādhvarin. Ed. N.S. Ramanuja Tatacharya. Kendriya Sanskrit Vidyapeetha Series 33. Tirupati, 1999.

Yogasūtra. With the commentaries of Vyāsa and Vācaspati Miśra. 3rd. ed. Ed. Jibananda Vidyasagara. Calcutta, 1940.

Binding Experiences for a First-Person Approach

Looking at Indian Ways of Thinking (*Darśana*) and Acting (*Nāṭya*) in the Context of Current Discussion on 'Consciousness'

SANGEETHA MENON

By following the current discussions on consciousness in the West, firstly, one gets to think that the understanding of consciousness is dependent on the understanding of if not brain, physical processes guided by a mechanism and having the capabilities for replicating the phenomenon *in vitro* with the help of controlled experiments; secondly there is not a consensual definition of the problem, method and the major goals of enquiry itself; and thirdly there is insufficient recognition of the very complexity and subjective nature of the phenomenon. All the three features have jointly contributed towards generating vast literature, dialogues and discussions about a variety of issues relating to consciousness, the primary being empirical research and medical possibilities, especially in the area of 'abnormalities'.

Introduction

Binding experience has been the single most issue in the center of focus in the last decade of discussions on 'consciousness' crossing disciplines: neurobiological, quantum mechanical, computational, theoretical, psychological etc. Though the details of what constitutes 'experience' differ from method and perspective, a consensus has emerged that (i) to explain 'consciousness' is to explain 'experience'; (ii) to explain experience is to explain its unity and binding nature. Following this preliminary consensus, however implicit it was, many discussions took place/are taking place from the first, second and

third-person perspectives, though the main stream discussion is still dominated to a greater extent by third-person approaches.

Given the complexity of 'experience' as a phenomenon for investigation, or as involved in our understanding, it is helpful to look at alternative views about what constitutes an 'experience'. I hope to do this with the help of instances from Indian epistemology and Indian dramaturgy. I will be looking at two different traditions of thinking and experiencing: Indian epistemology in the classical systems of Indian thought and Indian dramaturgy as dealt with in the classical text *Natya Sastra*. By doing this I hope to resurface the importance of 'experience' as lying in its nuances and juxtapose it with the revision as it is conceptualized now in the 'consciousness' discussions (which is reduced to third-person physical data, deprived of first-person intimacy, and also the depth and breadth of meaning). The attempt is to present the thesis that if consciousness cannot be understood without looking at 'experience', certainly experience cannot be reduced for convenient reductive (physical, psychoanalytic and cultural) methods of understanding but will have to be open for a variety of meanings validated from first-person perspectives. This will definitely take away the reductive scientific monopoly of explaining consciousness in a singular way, but will encourage scientific methods to reexamine the normative criteria for 'truth' and 'reality'.

The One Puzzle

I think there is an interesting and serious turn taking place in the current discussions on consciousness. This turn is based on and compelled by the intractable relationship of 'consciousness' with 'experience'. The nearest empirical idea for the unity and subjective nature of consciousness is 'experience'. Hence the scientific focus on 'experience'. The interesting part of the discussion is that though there is a recognition of experience as vital in the study of consciousness, the attempt itself is to strip 'experience' of the qualities which make it of experiential nature (unitary and subjective) and study it on the basis of empirical standards such as causal connections, neural influences, neural locations etc. I am not suggesting that brain research is not needed or even less important. Certainly, it is very significant in its own right. But if our guidelines and methods are not based on our basic premise to study consciousness (experience,

which is unitary and subjective) then certainly we cannot make a claim that brain studies apart from giving new knowledge about brain functions would also lead to a complete theory of consciousness. The puzzle in the current discussions on consciousness is that of the persistent conflict between epistemology and phenomenology.

If we look at the major semantic trends in the current discussions, the views which are discussed and debated no more fall into the classical division of reductionistic and non-reductionistic, or empirical and non-empirical approaches. However third-person the approach is, when it comes to the descriptive definition of consciousness, the ideas are based on qualitative features of consciousness. The discussions on empathy,[1] meaning,[2] meme,[3] and mirror neurons[4] are some instances. On the other side, the growing amount of discussions on meditation[5] and altered states of consciousness[6] give third-person references, however subjective the discussed experience is. A possible reason for this trend to interrelate and bridge first-person experience and third-person definition is the recognition of a distinct characteristic of 'consciousness', namely, that it is not completely defined by empirical standards and completely understood by first-person experience[7].

Self and Meaning

The extent of the meanings imputed to 'consciousness' most often crosses the empirical limits and sometimes even becomes diffused to qualitative experiential descriptions. The one major problem in consciousness studies is the semantics of 'consciousness'. Unfortunately this prominent meta-analysis of the discussions is dismissed in recent discussions[8]. It is very important that there is not only a well laid out definition for the problem but also a methodological consistency. This does not mean that even before the enquiry a complete theory of consciousness is anticipated. To have the semantics of consciousness given importance in the start itself means that the theory will not be drawn based on the limitations of the methods, but on the original contention about 'consciousness'.

What exactly are we trying to understand by the study of consciousness? The answers could range from neural functions to subjective experience. It is again interesting to see that the meanings we give for 'consciousness' are much larger conceptually than the strict semantic (in current discussions) definition of consciousness. This

is even clear at the starting point of the discussion when the immediate reference is to 'experience'.

It is in this context that I wish to propose the idea of 'self' as an alternative to the discussions of consciousness. The word 'self' is more comprehensive than the word 'consciousness' since it includes connotations at different levels of experience and also of the subjective identity which is important to understand the unity of experience.

Discussion about consciousness is discussion about experience. Discussion about experience is discussion about the 'self'. 'Experience' and 'self' certainly relate to something which is more than what is happening in the brain, more than abnormal conditions, more than ordinary conditions, more than transcendental states.[9]

Indian Thinking

There are two key ideas in classical Indian philosophical thinking which would strike the attention of any student. These are '*atma*' and '*darśana*'. These words perform a major double function, which is also the distinctive feature of the whole of Indian thinking, of combining epistemology and phenomenology. For this reason, '*atma*' could mean either the 'self' who is engaged in a particular act, or the self which is untouched by any act; *darśana* could mean discursive thought or intuitive thinking. The basic reason for such a foundational trend in the whole of Indian philosophical thinking goes beyond the felicity of a strict structural language (Sanskrit). It is an attempt not to break apart, and define, 'self' into identities based on the context; experience into ordinary and extraordinary; at the same time give thinking and understanding a depth which would be inclusive and open-ended but not divisive and hierarchical.

What constitutes darśana and what does not

Before I get to the details of the epistemology of *darśana*, I will briefly look at what constitutes *darśana*, and what does not. The word '*darśana*' connotes the philosophical enterprise to think and to delve with ideas so as to:

i) ascertain what is true knowledge,
ii) to understand new ideas, and,
iii) to understand the nature of the enquirer himself.

Jñāna is a complex concept in classical Indian thinking. It not only refers to logical and epistemological methods and answers but also states of mind which are important in the discussion about the primal nature of self. Hence, the discussions on *Jñāna* and *pramāṇa* are always interrelated to understanding ethical, axiological, aesthetic and spiritual issues. There is a constant attempt to reconcile and integrate different experiences, and the existence of contradic-tions so as to generate world views based on an understanding of life with answers for fundamental questions about self-identity, nature of world, creation, purpose of life, value systems etc. There is a wide-spread criticism that *darśana* does not have teleological value and does not extend its scope for change and modification. This perception might have been influenced by the complex method used by the schools using a variety of epistem-ological tools such as metaphors, imagery and stories, as equally valid, along with logical analysis, anticipating counterpositions and affiliation to definite theories of what constitutes right knowledge. Not strictly adhering to a definite pattern of enquiry could lead to the thinking that what is offered is a closed philosophical position to save the proponent and the follower of that particular tradition. What the *darśanakāras* are interested in is to give a new place to emerging ideas in the worldview and allow a new understanding crossing structural rigidity in thinking. Clearly, what is not the feature of *darśana* is an empirical haste to explain away things.

Epistemology of *Darśana*

Epistemological openness

Indian epistemology is constituted by complexes more than singular concepts. I will list a few such complexes, without going into the technical details, to demonstrate that Indian epistemology is an open-ended and integral enterprise.

Guidelines for discourse

Concepts and categories are vital to any kind of discourse. The school which perfected the art of discourse in Indian thinking was *Nyāya*. *Tarkasaṃgraha* which is the foundational text of logic and discourse is also the text followed, for that reason, by later schools in developing

their own theories. What makes *Tarkasaṃgraha* so very interesting and foundational is the way in which it defines and elucidates the necessary components for a discourse from both an epistemological and a subjective point of view. Both definition (*lakṣaṇa*) of an entity or idea, and the guidelines for discourse (*anubandha catuṣtaya*) are discussed with equal importance in the text. The meaning of the word *tarka* also is specific, in that it does not imply a pure logical analysis but a complex activity of discourse guided by strict definitions and goals so as to have "... a compendious elucidation of the nature of substance, qualities and such other ontological categories ...".[10]

There are sixteen *padārtha* which one studies in order to master *Nyāya* dialectics. A *padārtha* is defined as a "... knowable thing (*jñeya*) or as a validly cognizable thing (*pramēya*) or as a nameable or denotable thing which corresponds to a word (*abhidheya*)". These categories are means of knowledge (*pramana*), objects of valid knowledge (*prameya*), doubt (*saṃśaya*), purpose (*prayōjana*), instances (*dṛṣtānta*), established conclusions (*siddhānta*), members of syllogism (*avayava*), analysis (*tarka*), decisive knowledge (*nirṇaya*), arguing for truth (*vāda*), arguing constructively as well as destr-uctively for victory (*jalpa*), destructive argument (*vitaṇdā*), fallacious reasons (*hetvābhasa*), quibbling (*cāla*), specious and unavailing objections (*jāti*), and vulnerable standpoints (*nighrahasthāna*).[11] The discussion on *padārtha* is an elaborate one in classical thinking.

The concept of 'definition' (*laksana*) is another complex, which according to *Nyāya* tells what an entity 'is' by saying what it 'is not'. Definition is "... not merely an explication of the connotation of a term; but it is a proposition specifying the differentia or the differentiating feature of the species or the thing defined". *Lakṣaṇa* is defined as a specific feature (*asādharaṇa dharma*) which is free from the three faults of a definition such as over-applicability (*ativyāpti*), partial inapplicability (*avyāpti*) and total inapplicability (*asambhava*). A definition will be faulty by *ativyāpti* when it refers to certain qualities which is characteristic of the entity defined as well as of something not intended to be defined. A definition will be faulty by *avyāpti* when the definition does not refer to some of the characteristic features of the entity defined. A definition will be faulty by *asaṃbhava* when the definition refers to qualities which are totally non-characteristic of the entity defined.

Guidelines for teleology

Another important complex which is considered in almost all schools of Indian thinking is the notion of *anubandha catuṣṭaya* (four-fold preliminaries) though this is well specified as a part of dialectics in *Tarkasaṃgraha*. The four-fold preliminaries for any discourse is *viṣaya* (theme of discourse), *prayōjana* (major goal), *saṃbandha* (relation between the theme of discussion and the treatise), and *adhikārī* (for whom a discourse is designed).[12] The trend of specifying the objective and subjective guidelines of a discourse is also found in the foundational texts of *Vedānta* and *Mimāṃsā*. The starting verse of the text specifies the nature of enquiry such as for *brahman, dharma* etc.[13] The defining characteristic of a discourse clarifies any doubt which would ensue later in the discourse about what the discourse is guided by. The thematic specification of the discourse would also help the student to have a clear picture about what the discourse would not talk about or to what theme it would restrict to. Even if the theme of the discourse is given prior to entering into the discourse the discussion could at some point raise the question of teleology in the mind of the student. Hence the theme as well as the purpose of a discussion on such a theme is specified initially. Though it could be a meta-question outside the scope of the discourse it is essential also to anticipate at least to some extent the relation between the discourse and the theme of the discourse itself which would enable to understand how far the treatise or discourse is representative of the theme. The final and the most important preliminary factor for any discourse is to specify who is qualified to enter into such a discourse. This is a major rule for meta-discourse, which I think, is almost forgotten in the current discussions on a complex theme like 'consciousness'. The recognition of the aptitude of the person as playing a vital role in the success of discourse and understanding implies the subjective factor involved in epistemological enterprises. It also implies that understanding is always finally related to the basic aptitude of the student, which once again anticipates the essential relation between epistemology and phenomenology, knowledge of something and experience. One instance of expounding the nature of *adhikārī* could be seen in the primal text of *Advaita* '*Tattvabodha*' where Śankarācārya talks about '*sadhana catuṣtaya*'.[14]

Guidelines for validation

The issue of validation (*prāmāṇya*) is a very important complex extensively dealt with by the schools of Indian thinking. The discussion on validating knowledge ranges from theories of knowledge to theories of reality. The word '*pramāṇa*' etymologically means 'means of measurement' or 'that which produces knowledge'.[15] The concept of *pramana* though initially could be interpreted as a theory of knowledge, of ascertaining knowledge, its function will not be completely understood without taking into consideration two of the characteristic features of *pramana* as perceived by most of the classical schools of Indian thinking. These two characteristics '*abādhitatva*', of non-contradiction, and '*anadhigatatva*', of novelty, lays down the condition for validating knowledge.[16] A knowledge statement is out of validation to be true or false if there could be another knowledge statement which contradicts the claim of the previous statement. Being non-contradicted by another statement alone does not perform the role of validation. The characteristic of non-contradiction is to be also followed by the feature of novelty. Discovery of new knowledge is as important as ascertaining of it. Validation also has to look into the possibility of newness whether it is epistemological or ontological. The feature of 'novelty' implies once again the epistemological openness evident in Indian thought.

A major distinction in the Indian theories of knowledge is regarding the position on the origin (*utpatti*) and ascertainment of validity (*jñapti*). The validity of a cognition is decided, to some schools, by the presence of certain characteristics intrinsic to knowledge, and to some other schools, by the presence of certain characteristics extrinsic to knowledge. Following the same line of thinking, the two positions about invalidity of knowledge are that it is decided by extrinsic characteristics or intrinsic characteristics. Validity itself is ascertained to some schools by its very intrinsic nature (*svatapramanah*), and to some schools by its extrinsic nature (*paratapramanah*).

Two paradigms

There are two paradigms in the classical schools, inspite of the differences in their metaphysical and epistemological positions. These are i) what we see and experience, which is constituted by the

given and the immanent, ii) what we can see and experience which is constituted by the possibilities and the transcendent. It is within these two paradigms that the elaborate and detailed discussion on fundamental experiences such as pain and pleasure, sorrow and happiness, selfishness and selflessness, freedom and bondage, the given and the possible etc. takes place. *Darśana* is an attempt to bridge the seemingly two contradicting paradigms through an exploration of the self based on systematic discussions on i) theoretical, ii) experiential, and iii) transcendental issues.

What falls under theoretical issues

Theoretical problems are envisaged by the building of tools for thinking such as abstraction, generalization and conceptualization guided by the question of meaning, certainty and new knowledge. The factorization of 'new knowledge' in epistemology gives importance to intuitive thinking all through the discussion. A general division can be made of the theories the *darśanakāra*-s debate on, such as:

i) theory of what is given: which relates to ontological questions about the nature of the world, the nature of the self, the nature of life and death,

ii) theory of the what and how of knowledge which relates to epistemological questions about meaning and validity,

iii) theory of what is beyond the given (if any) which relates to metaphysical and teleological questions about the nature of God, the nature of ultimate causes, the nature of self and the nature of reality,

iv) theory of spiritual, mental and physical discipline which relates to questions about ethical issues, value systems, duty, responsibility, selfishness, transcendences and new perceptions about self-identity.

Experience of *Darśana*

Metaphysical openness

If we examine the classical schools of Indian thought, we find that though each school allows elaborate discussion on the epistemology of its philosophy, the foundational thought is metaphysical. But the metaphysical foundation is not be mistaken for dogmatic and closed

ideas. The metaphysical openness of ideas is evident from the fact that they are based on certain teleological assumptions. Discussions on the nature of (self) is juxtaposed with the physical (as in *Cārvāka* system), ethical or spiritual guidelines as in almost all schools. To understand the given nature of self and its transcendent possibilities the understanding of self is important. The key feature of such an understanding is that it is not an epistemological exegesis but a first-person phenomenological examination. The concept of *jñāna* is a complex concept and is not be merely translated as 'knowledge' as we understand it in popular fashion. The discussion on the given and the transcendent self (*jiva* and *ātmā*) is guided by the continous and rigorous distinguishing of the one from each other at every instance of experiencing. The conflict between the near and given nature of self, and the distant and transcendent nature of self forms the focus of attention for the *darśana*. The attempt of *darśana* is to solve the conflict in such a manner that the duals involved in it are integrated rather than segregated. The idea of liberation hence is not a singular event in time but a constant understanding and experiencing of the complexity of the contradiction of the given and the transcendent. The distinguishing of the *ātmā* and *anātmā* (the real nature and the given nature of self), *ātmā anātmā vyaparah*, is the singlemost exposition for which the rest of the epistemological, ethical and phenomenological theories are expounded. It is the metaphysical openness which is the hallmark of Indian thinking.

Spiritual and ontological openness

In recent discussions the word 'spiritual' has gained new meanings many of which emphasize the role of personal growth, ecological awareness, empathy, intersubjective transactions, emotional well being, efficiency in expressions, creative living. The distinct feature of the philosophical traditions of Indian thinking is its spiritual openness, by which I mean, not just a liberal philosophy, but the facility to integrate new experience and new understanding into an evolving scheme of ideas all leading and pointing to self-exploration. The ideal of spiritual living is given foremost importance than to moral and epistemological theories. It is not to say that the ethical guidelines and practices are less important in these traditions but to suggest that all such theories and discussions are addressed from a spiritual platform which discusses the nature of self and the world

of experience and the relationship between them. Liberation is the key concept however radically different the guidelines for it suggested by different schools are. Identity and self are the key problems addressed to with the help of metaphysical positions, epistemological theories and ethical guidelines. The breadth and length of discussions in *darśana* is interestingly just not different discussions on what exactly the nature of self is, but mutually reinforcing dialogues on the consensus view that all discussions are to be guided by the co-coordinating concept of 'self'. Invariably the discussions in *darśana* are discussions leading from the recognition of 'self' and 'identity' as larger categories for thinking. It could be because of this reason that epistemology (*tarka*) does not have the supremacy in deciding the course of events and validation, but does so only with the equal participation of reflective thinking (*vicāra*) in discourse. Analytical thinking could be delivering its goods only if it is accompanied by reflective (*vicāra*) and intuitive (*nidhidhyasana*) thinking.

What falls under experiential and transcendental issues

An interesting characteristic in the classical systems of Indian thinking is the overriding issue above all issues to connect and catapult from what would be considered as the given to what is possible. The conception about experience is not strictly what is caused by an extraneous factor/s but what could be possible by the distinctive and unique nature of the individual. Therefore, experience is not merely a theme for understanding based on its immediate context such as cause, or results, but a tool for further exploration of the self. The ordinariness and extra-ordinariness of an experience is understood more from the standpoint of the self than from the standpoint of what causes it. This trend also impels the understanding of the self along with the understanding of the object of experience. The object of experience, result of experience and the experiencer constitutes the triad of the complex phenomenon of experience, each one of which is significant in the understanding of each other.

The major experiential issues which are discussed in the classical schools are also interconnected with the major transcendental issues. Thus the experience and understanding of pain and pleasure are connected with guidelines for transcending pleasure and pain; experience and understanding of freedom and bondage are connected with the guidelines for transcending self-identities and rigid perceptions about the context; experience and understanding of different states of mind are connected with the guidelines for transcending words, verbal structures and attributed meanings.

Junctions and meeting points

The junctions and meeting points between the discussions on theoretical, experiential and transcendental issues are quite unique to *darśana*. For instance: ethical and spiritual discipline is necessary for new experiences and knowing the self differently; knowledge of self could change the way the nature of the given is understood; knowledge of self could reorient experience; knowledge of self could allow for new responses to the situation/context. What distinguishes the Indian way of thinking from what we today call the Western way of thinking is the curious connection present in *darśana* between theoretical, experiential and transcendental issues. It is also this distinguishing feature of Indian thinking which is often misappropriated as 'mystic' and 'other-worldly'. The important point missed here is that we fail to recognize that what interested Indian thinking was not the linearity and immediate conveniences through rigid structures of knowledge but an open-endedness where experience and reflection could together bring about a re-orientation of how we construe our self-identities and how we respond to the given.

The foundational issues, crossing the rigidity of being theoretical, experiential or transcendental, which are embedded in the *darśana* are i) about human mind, consciousness and experience, and ii) about self-identity. The guidelines for the exploration of these embedded issues are i) abstraction: to identify the unitary in the discrete, ii) placeability: to have an ontological meaning for any experience, its object and its experiencer, iii) practise: to have values and discipline as essential guidelines for self-exploration.

Indian Dramaturgy

The foundational text of Indian dramaturgy is '*Nāṭya Śāstra*' authored by Bharatamuni. The available form of the text is composed of 5600 verses coupled with prose though the original version is said to have

had more than 30,000 verses. It is a complete treatise on Indian dance, drama and music. The text has an exhaustive thematic structure since it deals with a complex conception of drama (*nāṭya*) constituted by what could be described as objective and subjective features. There is elaborate discussion, on one hand, on the characteristics of playhouses, different kinds of plays, different and complex gestures and movements, rules of prosody, metres and music, use of languages, style of characters, costumes and ornaments. On the other hand, there is discussion on emotions and mental states which are their causes, mutuality of emotions and mental states, rapport between actor and spectator, mental and physical nature of the actor and spectator, preliminary mystic rituals for effective representation and final goals of drama. At the same time there is a structural rigidity as to the epistemological structure, and openness about the subjective expression, relationship between the actor and the spectator, goals of drama etc.

The complexity of the text could be seen at three levels:

i) in addressing the representation of different kinds of characters (mostly mythical) with different states of minds through a joint participation of physical gesture and movements, mental states and emotions, ritualistic preliminaries, costumes, music and space configuration,
ii) in addressing the unique relationship between the actor and the spectator, of the actor invoking a specific state of emotion in the spectator's mind,
iii) in making possible a spontaneous and self-evolving nature of enjoyment for the audience in spite of the structured and specified composition.

The rigorous and specified rules of *nāṭya* together with an integral approach to emotions, first-person experience of the actor and the spectator make *Nāṭya Śāstra* an insightful treatise as well as what could be conceived of as belonging to a higher order of cognition and experience, namely a wholesome re-representation of human emotions through the complex act of external body (physical body gestures, costumes, music and plot) and the spiritual body (emotions, states of mind and unique relationship between the one who is presenting the re-representation and one who is enjoying it).

Epistemology of Nāṭya

The word '*nāṭya*' does not have a one-word English equivalent. Before we get to the meaning of '*nāṭya*' it is important to keep in mind the distinction between '*nāṭya*', *nṛtta* and *nṛtya* (*nāṭya nṛtya nṛtta vivekaḥ*)[17] which is the introductory theme discussed in *Nāṭya Śāstra*. *Nāṭya* is a combination of *nṛtya* (acting) and *nṛtta* (dance). *Nṛtya* is the visual and pantomimic representation of emotions and ideas. *Nṛtta* refers to movements of the body with gestures which are regulated by *tāla* (musical time interval). Though the text continues to give a complex definition of what constitutes *nāṭya* the categorical statement made about it is that *nāṭya* has primarily to do with *rasa*.[18] Later *nāṭya* is explained using two key ideas which are *abhinaya* and *bhāva*. *Nāṭya* means visual representation (*abhinaya*) in its fourfold forms such as using parts of the physical body (*āngika*), verbal utterances (*vācika*), costumes and ornaments (*āharya*) and physical signs of mental states (*sātvika*).[19]

Poise of expression

Abhinaya is defined as the expression through the actor of the meaning of the words of a literary (poetical) work with the help of *vibhāva* (emotions and states of mind physically represented) so as to invoke an uninterrupted flow of *rasa* (enjoyment) for the audience. The scope of *abhinaya* is extended beyond the rigidity of planned gestures and emotions by differentiating it to be of two distinct types: *lokadharmi* and *nāṭyadharmi*. *Lokadharmi* represents the objects and characters as they are portrayed in the mythical literature, and, *nāṭyadharmi* represents the objects and characters through suggestive movements from the setting of stage. An example of *nāṭyadharmi* is a suggestive movement of the eye or suggestive gesture by the hands to indicate something else through the imagination of the spectator.[20]

The fulfillment of *nāṭya* is achieved through the effective and joint performance of different kinds of *abhinaya* and *mudrā* (representation of objects, emotions and ideas through single hand and combined hand gestures), the theme of the play, music and the involvement of the spectators. The role of spectators is considered to be an active event and mutually influencing the performance of the actor in terms of the representation of feelings.

It is not directly relevant to this paper to describe the technical details of the themes of various chapters of the text. But it is necessary to keep in the background of the reading of the forthcoming discussion that the elaborate description in the text mainly follows two patterns:

i) discussion and detailed description of the different kinds of gestures of different parts of the body and their nuances; different kinds and features of plays and poetry; kinds of metres; characteristics of the actors, judges and spectators; use of languages; costumes and ornaments; and different kinds of musical instruments,

ii) discussion and description of *rasa* (emotions) and *bhāva* (mental states which produce emotions); the mental rapport between the actor and the spectator; the types of characters and mental and physical temperament suitable for their portrayal; the goals of drama and how they are fulfilled; and preliminary rituals and settings to invoke the conducive environment before the start of *nāṭya*.

The concept of *nāṭya* evolves in the text through the development of both the above patterns which I would like to describe as third-person and first-person approaches. The prescribed set of rules for *abhinaya* exists along with the spontaneity of the actor in representing the structured, and in evoking the *rasa* in the spectator. The visual and the character-oriented together with the subjective and self-oriented produce the aesthetic experience which could be further described as a spiritual experience. The act of representation, the preliminary settings and rituals etc. connected with the cosmogony that the physical world is the *āngika abhinaya* of Śiva, his *vācika abhinaya* is the world of language, his *āharya abhinaya* consists of the universe and his *sātvika abhinaya* is ultimate happiness itself.[21] The complexity of representing human emotions and at the same time invoking empathy in the spectator is brought out through *natya* in a comprehensive manner using a rigorous epistemology and first-person experience for both the actor and the spectator.

Experience of Nāṭya

Nāṭya is though presented following a structured design about it through the portrayal of characters, the primary nature of *nāṭya* is experiential and first-person-oriented. This is evidenced by the detailed discussion on *rasa, bhāva* and *prēksakatva*. The word *nāṭya* has its origin from the root '*naṭ*' which means 'to act'. *Naṭa* is one who performs the act through different styles of *abhinaya*. *Nāṭya* is the art of '*naṭa*'.[22] The importance given to *nāṭya* as a dramatic art has its origins in the 'act' itself of the actor. It is the *naṭa* who is responsible for *nāṭya* and not vice versa. This is a significant feature since it emphasises the first-person oriented approach to a complex event such as *nāṭya*.

Tasting the flavour

Rasa is a complex concept which is the central idea on which the experience of *naṭya* is founded. The word *rasa* is variously translated as 'relish', 'enjoyment'[23] and related to mean the object of relish or relish itself. According to Bharatamuni *rasa* emerges out of the combination of three basic components such as *vibhāva, anubhava* and *vyabhicari*. They are also the (*kāraṇa, kārya* and *sahakāri*) determinant, consequent and auxiliary conditions of *rasa*. All three taken together is called the *sthāyibhāva* which is directly responsible for the production of *rasa*. *Bhāva* is that which makes something happen.[24] In *Natya Sastra* '*bhāva*' is used as a technical word to relate to the mental states as responsible for producing *rasa* for the spectator through a combination of kinds of (*abhinaya*) gestures. Whether *rasa* is produced through *bhāva* or vice versa or whether they are mutually influenced is a debate which is prominent in the literature on *Nāṭya Śāstra* by various commentators. For the discussion in this paper, I will deal only with the detailed presentation of kinds of *rasa* and *bhava*, one instance of *abhinaya* which is that of eyes (*dṛṣti*), and nature of effectiveness of *nāṭya* (*nāṭya siddhi nirupaṇa*), to show the importance given to the nuances and details of mental states, basic nature of experience and their physical representations, with an attempt to give a third-person account of first-person experience.

Bharata enlists eight *rasa*-s as the primary *rasa*-s[25] and a total of fortynine *bhāva*-s which are classified as *sthāyibhāva* (eight in number), *vyābhicaribhāva* (thirtythree in number) and *sātvikabhāva* (eight in number). This classification refers to an evolution of mental states from its intense and pure states (*sthāyī*), to manifestation of the pure states in feelings and leading them to *rasa* (*vyābhicāri*), and to their physical signs (*sātvika*). Another classification is of the cause and effect of *bhāva* such as *vibhāva* and *anubhava* respectively.

The *sthāyibhāva* of *śoka* is produced by the *vibhāva* such as separation from the beloved, loss of dear ones and assets etc. *Śoka* is represented by the *anubhava* such as tears, deep sighs etc. It is the *sthāyibhāva* which plays the key role in creating the rapport between the actor and spectator through the production of *rasa*.

Bharata enlists eight fundamental *sthāyibhāva*, thirty three *vyabhicāribhāva*, eight *sātvikabhāva* and eight *rasa*[26] according to the *sthāyibhāva*. These are:[27]

RASA	STHĀYIBHĀVA
Rati (happiness)	**Sṛngāra** (charm)
Hāsa (laughter)	**Hāsya** (humour)
Śoka (mental pain)	**Karuṇā** (compassion)
Krodha (anger)	**Rudra** (fury)
Utsāha (enthusiasm)	**Vīra** (heroic)
Bhaya (fear)	**Bhayanaka** (terrifying)
Jugupsā (disgust)	**Bhibhatsa** (despicable)
Vismaya (amazement)	**Adbhuta** (surprise)

The thirty three *vyabhicaribhava* (all these are given their corresponding *vibhava* and *anubhava* in the *Natya Sastra*) are:

VYABHICĀRIBHĀVA
Nirvéda (disinterest)
Glāni (tiredness)
Śanka (apprehension)
Asuya (insecurity)
Mada (intoxication)[28]
Śrama (exhaustion)
Álasya (lethargy)
Dainya (pity)
Cintā (anxiety)
Moha (delusion)
Smṛti (recollection)
Dhṛti (steadfastedness)
Vṛidā (shame)
Capalatā (impulsiveness)
Harṣa (sudden delight)
Avēga (excitement)
Jaḍatā (stupor)

Garva (arrogance)
Viṣāda (depression)
Autsukya (longingness)
Nidrā (sleep)
Apasmāra (epilepsy)
Supta (dreaming)
Vibōdha (awakening)
Amarṣa (restrained anger)
Avahitthā (deception)
Ugrata (ferocious)
Mati (analytic understanding)
Vyādhi (ailment)
Unmāda (temporary loss of sanity)
Maraṇa (death)
Trāsa (panic)
Vitarka (argumentativeness)

The eight *satvikabhava* are:

SĀTVIKABHĀVA

Stambha (paralysis)
Svéta (perspiration)
Romānca (horripilation)
Svarabhēta (change in the tone of voice)
Vépathu (tremble)
Vaivarṇya (change in the color of face)
Aśru (tearful)
Pralaya (fainting)

Through the corners of the eyes

Expression through the physical body, *āngika abhinaya*, is further classified into that falling in three types such as i) *śariraja* (bodily), ii) *mukhaja* (facial), and iii) *ceṣtakṛta* (through movements). There is another division of *anga* and *upānga*. *Anga* constitutes the *abhinaya* through head, hand, chest, sides, hips and feet. *Upānga* constitute *abhinaya* through eyes, eyebrows, nose, lips, cheeks and chin. The one instance of *abhinaya* through *upanga* I will list here in detail is that of *dṛṣti* (glances) since these are considered to be more visually representative of the *rasa*.

What I wish to imply through this listing is, the analysis and observation given to the detailed study of empirical features of inner mental states and feelings belonging to another person as represented by the actor. The two levels of third-person reporting and first-person experience are interesting to note at this point. Through the *bhāvabhinaya* the actor represents the feelings of a person with a particular state of mind through the larger setting of stage, space, costumes and gesture (first instance of third-person reporting), and all the while undergoes the same state of mind so that the corresponding *rasa* is conveyed to the spectator (second level of third-person reporting). The enactment of the feelings are based on an understanding of the *bhāva* (pure states of mind) and identifying with them (second level of first-person experience) which was earlier experienced by another person (first level of first-person experience).

The glances which total thirty six in number are of two kinds: i) *rasadṛṣti*, representative of the kind of *rasa* (relish), and ii) *bhavadṛṣti*, representative of the kind of *bhāva* (pure state of mind). Apart from the *dṛṣti* there is a detailed description of the kinds of movements of the pupils, eyelids, and that of eyebrows.[29]

The eight *rasadṛṣti*[30] are:

RASADṚṢTI NATURE OF GLANCE	RASA
Kānta (loving glance) eyebrows moved, glance through the sides of the eyes, and eyes with the intense look of as if drinking the object	Sṛngāra
Hāsya (humorous glance) eyelids are contracted one after the other, wandering pupils	Hāsya
Karuṇā (compassionate glance) upper eyelid droops down with tears, wandering pupils, nose ends intense	Karuṇā
Raudri (ferocious glance) both eyelids tremble, still pupils, red and dry eys, strained eyebrows	Raudra
Vīra (heroic glance) steady pupils, fully opened and glowing eyes, ends of the eyes contracted	Vīra
Bhayānaka (terrifying glance) raised and motionless eyelids, restless pupils	Bhayānaka
Bibhatsa (disgusting glance) eyelids come together with restlessness, unsteady pupils, eyeballs at the corners of the eye[31]	Bibhatsa
Adbhuta (surprising glance) moist eyes, pupils go in and out alternately, eyelashes slightly contracted, bright corners of the eyes[32]	Adhbuta

The twentyeight *bhavadṛṣti* are:

STHĀYIBHĀVA DṚṢTI NATURE OF GLANCE

Snigdha (tender glance) fully opened eyes, eyebrows held up, pupils in the corners of the eye
Hrsta (joyous glance) slightly contracted pupils which are restless, eyelids close alternately
Dina (piteous glance) drooping upper eyelids, restrained movement of pupils, with tears
Kruddha (glance with anger) motionless eyelids, dry eyes, agitated pupils, bent eyebrows
Drpta (glance with pride) fully opened eyes and still pupils
Bhayānvita (glance with fear) fully opened eyes, eyeballs standing out and agitated[47]
Jugupsita (glance with disgust) contracted eyelids, look away from the object, indefinite look
Vismita fully open eyelids, steady look to a distance, pupils held up[33]

VYABHICāRIBHAVADṚṢTI[34] NATURE OF GLANCE

Śūnya (vacant look) steady eyes, but not clear, look vacantly without an object[35]
Malina (depressed look) pupils directed away from the object, eyelids slightly closed, clear corners of the eyes, throbbing eyelashes

Śrānta (tired look) pupils directed to a short distance, moist eyes, tired eyeballs, slightly contracted corners of the eyes
Lajjita (glance with shyness) eyelashes come together, tired pupils, drooping upper eyelids
Śankita (glance with suspicion) eyes are alternately steady and restless, turn towards the sides outward and upward, alternate intense looks and looking away
Mukula (fully closed look) united and throbbing eyelashes, resting pupils
Ardhamukula (half opened look) half opened eyes and slightly throbbing, half closed eyelids
Glana (languid look) deeply sunk pupils, move very slowly, eyelashes, eyebrows and eyelids appear like that of a blind person
Jimha (looking distrustfully) slightly contracted eyelids, tired and concealed pupils, look slowly
Kuncita (contracted look) eyelids and eyelashes are slightly contracted, pupils are well contracted
Vitarikta (look of indecision) raised eyelids, flushed and downward pupils
Abhitapta (extreme painful look) gentle movement of eyeballs, upward and downward moving eyelids, all the parts of eye indicate extreme pain
Viṣaṇṇa (grievous look) corners of the eye are sunk, eyelids wide apart and open and close frequently, motionless pupils
Lalitā (charming look) corners of the eyes contracted, eyebrows go up and down, and sweet look
Ākekārā (half closed) eyelids at the corner of the eye are slightly contracted, half closed look, pupils are repeatedly turned
Vikosa (wide open look) fully open eyelids and never close, unsteady pupils
Vibhrānta (distracted look) occasional disturbed and undisturbed look, moist and wide open eyes, moving pupils
Vipluta (floating look) steady and drooping eyelids in succession
Trāsta (fearful look) extremely unsteady pupils, eyelids quickly moving up and down
Madira (intoxicated look) Is classified into three kinds such as the early (taruna), middle (madhyama) and extreme (adhama) stages
Taruna corners of the eye are wide and the rest of the eye is contracted, pupil move about in a circle

Madhyama slightly contracted eyelids, unsteady pupils
Adhama pupils move downward and eyelids are almost closed[36]

Twilight space of the real and the virtual

Bhāva and *rasa* are the two key concepts according to Bharata. Though he considers *nāṭya* to be effective as a result of performance, he makes a detailed analysis of how the performance which is not realistic and identical with the world of reality for both the actor and spectator are made real in a virtual manner. The *sthāyibhāva* enlisted are pure states which can exist and together with *vyābhicari-bhāva* and *sātvikabhāva* produce the necessary *anubhava*, only if the actor could form an identity with the *sthāyibhava*. The performance of the character is dependent on this preliminary identification of the actor with the state of mind of the role which is portrayed. The *sthāyibhāva* are made to exist (*bhāvayanti iti bhāvah*) by the actor so that it would be produced in the mind of the spectator (*bhāvanti iti bhāvah*) to produce the related *rasa*. The *sthāyibhāva* together with *rasa* could be considered as causing self-transcendence for the actor as well as the spectator, and complex cognitive structures for both the actor and the spectator to be in communion in a space where both transcends their self-identity. It is an enactment (by the actor through *nāṭya*) which is more spontaneous than the simple mimicry of an event or object. It is at the same time physical and trans-cendental.

One of the unique features of *nāṭya* is that the epistemological and the experiential, the theory and technique are co-ordinated to form a mutually benefiting factor of the whole. Though the source of the following text is not authentically traced, it is said in both *Nāṭya Śāstra* and later in *Nāṭyadarpaṇa*, and is also popular to be the synoptic definition of *nāṭya*, that 'the body should follow the tune, the hands must explain the meaning, eyes must speak the emotion and the feet must beat the time-measure; where the hands go there should go the eye, because where the eye goes there the mind goes with it, where the mind goes there follows the mental state, where the mental state is, there is the feeling'.[37] These two verses represent the coordinated physical, mental and transcendental nature of *naṭya*. Equal importance is given to detailed and specific physical and mental factors involved, and each of their transcendence is specified, at the same time, to broaden the scope of experience both for the actor and the spectator.

The metanarrative

That *nāṭya* is taken a wholesome event is evident from the fact that apart from the detailed account of the content of *nāṭya* Bharata also devotes separate chapters[38] for looking at the effectiveness of *nāṭya* (*nāṭya siddhi nirupaṇa*), detailed description of the nature of actors, judges and spectators, and the goals of *nāṭya*. There is even a mention about the seating arrangement to be followed.[39] It might be keeping the complex nature of *nāṭya* that Bharata enlists for male and female characters: i) three classes of personality (*uttama, adhama* and *madhyama prakrti*), and , ii) the kinds of roles they could play.[40]

The spectator of *nāṭya* is not a passive recipient, but a *prēkṣaka*, 'one who views in a unique manner'. It is evident that Bharata included the active and important participation of the spectator for *nāṭya* to be a successful enterprise. There is a list of physical representations of the responses of the spectator to *nāṭya* by making certain words,[41] sounds[42] and physical and facial expressions.[43] There is also a description about who is the genuine spectator (*prēkṣaka*). He who is ". . . one who has unruffled senses, is pure, clever in discussing and weighing pros and consciousness, devoid of faults and fond of merits. He who attains gladness on seeing another glad, sorrow on seeing another sorry and experiences wretchedness on seeing the wretchedness of another is considered fit to be a spectator . . .".[44] It is also said that all these qualities may not be present in one single individual, but the different individuals as spectators could have them and together make an effective appreciation of the *nāṭya*.

The goals of *nāṭya* pertain to both objective and subjective features. Through the composite of external and physical enactment, and subjective states of mind and feelings representative of them, what is achieved for the i) actor and ii) spectator are: *For the spectator*, in the secondary level an appreciation of the characters and the theme, and in the primary level a temporary detachment with his/her self-identity is experienced. *For the actor*, in the primary level it is the complex task of representing a character, an idea or a nuance of a particular feeling through *abhinaya* and producing the corresponding *rasa* for the *prēkṣaka*. In the secondary level a temporary detachment from his/her self-identity and identity with the particular character's self as a whole and various mental states which the character would have in the story narrated. The transcendence experienced by the actor is both transphysical and transmental since there is the combined use of body and mind. The transcendence experienced by the spectator is transmental.

And finally it is experience and transcendence

For both the actor and the spectator it is a complex experience since there is the co-existence of his/her own dominant and real self-identity, and the identity with the mental states of the character portrayed. It is the co-existence, of the real self-identities of the actor and the spectator, and the identities with 'another-self', which determines the effectiveness of *nāṭya*. The interesting and intriguing feature is the existence of a contradiction. For the effective transference of a particular *bhāva* to the spectator the actor has to have an identity formed with it transcending the artificiality of enacting it. At the same time the actor has to be detached from any specific *bhāva* of the character since what he/she is primarily concerned with is the narrating of the story. The actor has to play the twin role of 'being the character portrayed' and also the narrator of the story. It is this twin and contradictory role played by the actor which enables the spectator to have the experience of *rasa* which also involves an interesting contradiction. Unless the spectator can be one with the mental state of the character portrayed he/she will not be able to appreciate the story and the specific nuance. At the same time unless a continous detachment is maintained he/she will not be able to integrate the experience of that nuance in relation to his/her self-identity.

Re-placing Consciousness (in Indian Thought)

By presenting two different instances of epistemology and experience from *darśana* and *natya*, what I wish to suggest is that:

i) contrary to the much popular and published view that Indian philosophy is 'other worldly', there is detailed and careful presentation of what could be considered as the two primary signs of consciousness namely (a) generation of meaning and its validation (b) intensity of experience and broadening its scope through its own transcendence,

ii) the discussion on 'consciousness' in Indian thought is not a word-oriented (namely 'consciousness') but an experience-oriented task which looks at empirical, experiential, epistemological and

teleological facets of consciousness.

Present approaches to understanding of consciousness though to some degree gives importance to epistemology and to some degree gives importance to first-person experience what is missing is an attempt to resist untimely classification of events and meanings of 'consciousness' under empirical/medical/ordinary and transcendental/psychoanalytic/mystical groups and their segregated and non-dynamic explanations. To be in the context of particular experiences, and to integrate them to a transcendence which will least look unfamiliar and 'other-worldly', cannot be the result of classificatory understanding or solipsistic transcendental experiences alone.

First and foremost we need to recognize 'consciousness' as a complex phenomenon and thereby dissuade ourselves from secluded and segregated analysis. Complexity of consciousness more and more looks like the delicate togetherness of understanding and being. The understanding of 'consciousness' is more an understanding of its ontology, which needs the focus of epistemology to be shifted from normal and ordinary experiences, abnormal and transcendental experiences, to the holistic definition of the problem, method, and goals of enquiry. This would facilitate breaking 'habitual' ways of event or object oriented analysis by experience or first-person oriented understanding. The categories of thinking formed by the analyst and his/her worldview will be specific as well as potential for widening the scope of understanding.

The two questions which are important, if we are 'really' interested in understanding consciousness are: (i) What are we really looking at?, and (ii) What do we really want to look at? Our notions about 'real', 'truth' and 'self' are to be continuously questioned, but at the same time, integrated with personal growth, values, spiritual understanding and self-exploration.

Acknowledgment

My Pranam to Swami Bodhananda for his Blessings without which the completion of this paper would not have been possible. Thanks to Leela Ramanathan (President, Karnataka Nrtakala Parishat) for discussions and comments. Thanks also to Rajiv Malhotra for encouraging me to write this paper.

NOTES AND REFERENCES

1. Evan Thompson (2001). "Empathy and Consciousness". *Journal of Consciousness Studies.* 8 (5-7) 1.
2. Patricia Churchland (1986). *Neurophilosophy: Toward a Unified Science of the Mind Brain.* Cambridge: Bradford Books.
3. Susan Blackmore, (1999). *The Meme Machine.* Oxford University Press.
4. V. Gallese (2001)."The shared manifold hypothesis". *Journal of Consciousness Studies.* 8(5-7) 46.
5. Jensine Andresen (2000). "Meditation meets behavioural medicine: The story of experimental research on meditation". *Journal of Consciousness Studies.* 7(11-12) 17-73.
6. Charles Tart (1992). Ed. *Altered States of Consciousness.* 3rd ed. Harper Collins.
7. Menon, Sangeetha (2001). "Binding experiences: looking at the contributions of Adi Sankaracarya, Tuncettu Ezuttacchan and Sri Narayana Guru in the context of recent discussions on consciousness studies". URL: www.infinityfoundation.com/mandala/i_pr/i_pr_ker_menon_frameset.htm Paper presented at the 12th annual national conference of National Academy of Psychology.Kollam.October 22-24, 2001.
8. Michael V. Antony (2001). "Is 'consciousness' ambiguous?". *Journal of Consciousness Studies.* 8(5-7) 19.
9. Sangeetha Menon (2001). "Towards a Sankarite approach to consciousness studies". *Journal of Indian Council of Philosophical Research.* XVIII(1) 95-111.
10. S. Kuppuswami Sastri, (1932). *A Primer of Indian Logic: According to Annambhatta's Tarkasamgraha.* Mylapore: Madras Law Journal Press.p.4
11. Ibid., p. 6
12. Ibid., p. 3
13. The *Brahmasūtra* begins with the sūtra '*adhato brahma jijnasa*' and *Mimamsasutra* begins with the sūtra '*adhato dharma jijnasa*'.
14. The introductory theme in *Tattvabodha* is '*sadhanacatustaya*' which talks about the fourfold qualifications needed for a student interested in the enquiry of moksa. The fourfold qualifications are: (i) *nitya anitya vastu viveka* (discriminatory understanding of the real and the unreal) (ii) *iha amutra artha phala bhoga viraga* (dispassion towards the objects of pleasure) (iii) *sama adi satka sampatti* (observance of the seven values) (iv) *mumuksutvam ca iti* (earnest desire for liberation).
15. prama karanam pramanam in S. Kuppuswami Sastri (1932). *A Primer of Indian Logic: According to Annambhatta's Tarkasamgraha.* Mylapore: Madras Law Journal Press, p.189.
16. M. Hiriyanna (1975). *Indian conception of values.* Mysore: Kavyalaya Publishers, p. 4
17. *caturtha abhinaya upetam laksanavrttito buddaih*
nartanam natyamityuktam sa tvatrabhinayo bhavet
K. Vasudeva Sastri, and G. Nagaraja Rao (1990), Ed. & tr. *Natya Sastra Samgraha.*3rd ed.vol.1. Tanjavaur: Saraswati Mahal Library, p. 20.
henceforth denoted as Natya Sastra Samgraha
18. *natyasabdo rase mukhyo rasa abhivyaktikaranam*
Samgraha: p. 20

19. *caturthabhinaya tatra angiko angaidarsito matah l*
vaca viracitah kavyanatakadistu vacikah ll
aharyo harakeyurakiritadivibhusanam l
satvikah satvikairbhavaih bhavukena vibhavitah ll
Nāṭya Śāstra Samgraha: p. 20
20. In *vacika abhinaya* speech is *lokadharmi* and singing is *nāṭyadharmi*. In *aharya abhinaya* wearing of ornaments is *lokadharmi* and suggesting objects by mere gestures is *nāṭyadharmi*. In *satvika abhinaya* shedding tears is *lokadharmi* and suggesting tears by gesture is *nāṭyadharmi*. *Nāṭya Śāstra Samgraha*: p.25
21. *angikam bhuvanam yasya vacikam sarva vangmayam l*
aharyam candratarati tam namah satvikam sivam ll
Natya Sastra Samgraha: p.1
22. *The Natya Sastra of Bharatamuni: Translated into English by a board of scholars.* (2000). Delhi: Sri Satguru Publications. 2nd revised ed.p.528
henceforth denoted as Natya Sastra
23. *rasa iti kah padārthah*
ucyate asvadhyatvat
G.H. Tarlekar (1991). *Studies in the Natya Sastra: With special reference to the Sanskrit drama in performance.* Delhi: Motilal Banarsidass Publishers pvt.ltd. 2nd revised ed.p. 54
henceforth denoted as *Studies in Natya Sastra*
24. *bhavayanti iti bhavah, Natya Sastra*: p.86
25. *srngara hasya karuna raudra vira bhayanakah l*
bibhatsa adbuta samjnau cetyastau natye rasah smrtah ll
Studies in Natya Sastra: p. 56.
Bharata talks only about eight *rasa*. It is Abhinavagupta who introduced the ninth *śānta rasa*, and also as the most important *rasa*, which was essential to portray the unique spiritual nature of Buddha.
Studies in Natya Sastra: p. 60.
26. Translations of the terms are by the author.
27. *Natya Sastra*: pp. 86-113.
28. Three kinds of mada are mentioned according to their instensity.
See *Natya Sastra*: p. 95.
29. *Natya Sastra*: pp. 123-126.
30. *Natya Sastra Samgraha*: pp. 483-491, *Natya Sastra*: pp.118-119
31. Ibid., pp. 483, 489.
32. Ibid., p. 450.
33. Ibid., p. 492-499. *Natya Sastra*: pp. 119-120.
34. Ibid., p. 500-524. *Natya Sastra*: pp.120-122
35. Ibid., p. 50.
36. Ibid., pp. 501, 521.
37. *angena alambayet gitam hastena artham pradarsayet l*
netrabhyam darsayet bhavam padabhyam talamacaret ll
yato hasta tato drstih yato drstih tato manah l
yato manah tato bhavah yato bhavah tato rasah ll
Natya SastraSamgraha, p. 31.
38. *Natya Sastra*, Chapters.27, 34, 35.
39. *Natya Sastra*, Chapter.27, p. 381

40. See *Natya Sastra*, Chapters. 34, 35, pp. 514-530 for a detailed description of the classes and kinds of role.
41. See *Natya Sastra*, p. 376 . Words like '*kastam*' for the pathetic feelings portrayed.
42. See *Natya Sastra*, p. 376. Words like '*aho*' for implying 'how wonderful' the portrayal is.
43. See *Natya Sastra*, p. 375. Appreciation of humour is implied with smiles and laughter; appreciation of joy is expressed through horripilation.
44. *Natya Sastra*, p. 380.

The Nature of Human Motivation

INDRANI SANYAL

One of the perennial concerns for human beings is not with things or facts or events as they are, but also with what they should do or what they ought to be. In spite of society being composed of a plurality of persons, each with his own aim, distinctive interests and the conception of good, society endorses certain principles for beneficence, justice, non-malevolence, respect for the autonomy of the person and these principles have become some of the paradigms for ought statements. One should love one's neighbours, one should not cause injury to other fellow beings, one should not steal others' property, one should not kill animals, one should protect one's environment—are just instances of some of the frequently pronounced ought statements. With the rapid advancement in science and technology newer and newer ought and ought-not questions are coming to the forefront. Some philosophers have discovered a close link among these vexing ought questions, their being put to effect, and consciousness, and have offered various arguments in defense of their claim. To act according to a certain ought one has to be motivated to act in that way. Theories to account for the nature of motivation are numerous. The sociobiologist would explain motivation wholly in terms of genetic inheritance as providing the basic programming to facilitate survival and reproduction. The instinct theorists would consider the happening of behaviour as a preprogrammed part of the organism. The Freudian would come up with his theory of *eros* and *thanatos*, the life instinct and the death instinct to account for our conscious life. None of these approaches would agree with the account of motivation as given by Indian philosophers of theistic outlook. To explain why a person moves to perform an action and why one refrains from performing an action is to explain the what and the why of voluntary action. To say that one ought to act in a certain way is definitely not to compel the person to perform the act concerned under duress or otherwise. The reference to the immemorial problem of free will *vs* determinism

can hardly be altogether avoided in any moral discussion. If all of a man's actions are determined and if his will is not free, how can he be held responsible for his actions? And if he cannot be held reponsible for them, what justification can there be for reward or punishment, praise or blame ? Does it not then make the study of ethics a futile venture?

The materialistic determinists do not demur from assuming that all the thoughts, values, volitions, decisions, acts are the products of physical, chemical or physiological processes going on in the human body. In such a view the human mind or will can originate nothing. It just transforms outward pressures and forces or inward chemical changes into ideas or acts, or produces the illusion of volition or free will. If a dynamo can automatically transform motion into electricity or if an engine automatically transforms steam, electicity or gasoline into motion why can't gross physical body generate will? These hardcore materialists do not desist from announcing that so-called will is merely the name for an automatic and predictable process. The Carvakas or the Indian materialists also argue in the similar vein. This is indeed one alternative model from which the nature of human volition and motivation has been understood. In this paper my objective is to present its counter model that shuns altogether the panphysicalism of the previous materialistic model. The Nyāya Vaisesika thinkers and the Prabhākara Mimāmsakas in India have elaborate discussions on the topic of will or volition. The diverse theistic schools in Indian philosophy generally agreed in believing in the authority of the Vedic texts, believing in an imperishable self, other than the perishable body, believing in a certain realism about the external body, i.e., the gross physical objects such as pots and jars, stones and mountains.

This paper will give an analysis of volition as already mentioned from the Nyāya Vaisesika point of view for this constitutes one of the major psychological bases of theistic Indian philosophy. There are other aspects of moral consciousness which are not discussed at present. Each of us in a society has two important roles to play—either of an agent who performs the action or that of a patient, one who is affected by the performance of that action. A third point of view, that is the role of a disinterested or impartial observer, may also emerge at times. To frame or develop any workable or acceptable theory about human motivation one needs to put oneself imaginatively in the role of each of the persons that a given ought statement

would affect in one way or the other. However, in the present paper the whole analysis is carried out from the first person perspective of the motivated agent who is the doer of the voluntary action.

From the analysis as given by these philosophers it will be seen that volition, as understood by these thinkers, is determined by conscious choice and also by the consciousness that the choice has been free i.e. determined by my own will. However, no one should harbour the wrong idea that everything that is materialistic is Western and everything non-materialistic or rather spiritualistic is Indian. The Nyāya Vaisesika model of volition is just one of the models to explain and understand human volition in a non-materialistic pattern. Aristotle also made a careful analysis of that aspect of human nature that is involved in acting and accepting moral responsibility. To attribute responsibility to a moral agent presupposes, according to Aristotle, the undertaking of moral actions voluntarily. In the Aristotelian Ethics we find clear-cut statements about two distinct conditions under which human actions are involuntary. An agent is not responsible for actions that are produced by some external force. Thus, if somebody grabs my arm and uses it to strike a third person, I cannot reasonably be blamed. Thus, if someone at gunpoint made me kill a third person then I am not to be morally blamed for what I have done. To the list of involuntary actions, according to Aristotle, is to be included actions performed out of ignorance. The main lesson that is conveyed by Aristotle is that moral actions are within our capacity to perform or avoid.

In Indian philosophy, the thinkers like Prasastapada, Visvanath and many others provide elaborate accounts of volition. Salikhanath Misra in his *Prakaranapanicka* also elaborates the position of the Prabhākara Mimāmsakas on this aspect of volition. The discussion of volition proper as distinguished from automatic or reflex actions will throw some light on the nature of the human individual as person. Gotama, the founder of Nyāya, proved the existence of self by the feature among other things of having volition. The expression 'Pravrtti' has been variously translated, as impulse, motive, will or volition. In this view, motive is an internal state. As human beings we perform various activities. Broadly speaking, all activities performed by human beings are of two sorts: in one case the desire or aversion of the actor who performs the act, in the performance or the non-performance of it, has some decisive role to play; but in another sort of action the agent's desire or aversion are non-starters. Various types of organic action and automatic action have their

suitable effects and also serve some specific ends for the organisms. But these actions do not fall within the domain of actions that presuppose either the desire or aversion of the agent concerned. Now voluntary actions are there to serve certain ends viz, the selection of the good and the rejection of the bad. Two things clearly emerge from the Nyāya Vaisesika account:

a) a mental activity followed by
b) a bodily motion.

The Nyāya Vaisesikas give an account of human activity from the standpoint of myself as the agent or the actor in the situation, this may be called the description of human activity from the first person perspective. It is as if the subject looks into himself and tries to find out what the factors are that motivate him. In the first place, a person is motivated to act if he has a desire to perform that act. Broadly speaking, desire is defined as the yearning for the unattained. Desires are sometimes self-centred, egoistic and sometimes all-involving, altruistic. In the Nyāya literature a long list of various desires are given, ranging from sexual desire, appetite for food and drink, desire to remove the suffering of others, desire to renounce the world, desire to deceive others, and finally desire for action and various other forms of desire arising from different activities. In this context the discussion of desire needs some further clarification. There is also another sort of classification of desire according to whether the desire is for an end or consequence that is good or whether it relates itself to the means or the method for achieving a certain end. In the case of volition it is the desire of the second kind that is the desire for the means for attaining the end.

This desire for action has as one of its originating conditions, as assumed by the Naiyāyikas, the connection of soul with mind. This shows that the existence of self and mind are two basic pre-suppositions to account for volition from the Nyāya Vaisesikas perspective. The desire to perform an action is an extraordinary type of desire as it is a desire to perform an act which is cognised or believed to be such as capable of being accomplished by the will. The awareness or belief that a certain act can be performed implies some sort of faith in one's capability. To desire to perform an act, the act must not be such, which is not in the competency of the agent to accomplish. As an agent, one cannot desire to perform an act for which one's competency is lacking. The consciousness of the

act's being accomplished by me is a necessary condition for the desire to act. It is a desire or wish to do that which I reoognise I am to do and therefore, it is in my capacity to do. The Naiyāyikas point out that the mere belief that some act is fit to be done by me is not sufficient to arouse desire in me to perform that act. The Naiyāyikas elaborate the following conditions to account for the desire to perform an action:

1) consciousness of an act as fit to be performed by me.
2) the belief that the performance of this act is conducive to my benefit.
3) the belief that the benefit is unaccompanied by any stronger evil.

It has been made amply clear by the Naiyāyikas that if there is no consciousness that the performance of the act is beneficial for me, then there is no volition on my part towards that act. The Naiyāyikas are consequentialists and they account for the motivation in terms of a preconceived beneficial end or consequence. For the Carvakas the good is either pleasure of the senses or body or freedom from any bodily suffering. In this connection it may be asked, how can we justify a motive to suicide from the perspective of the Naiyāyikas? The agent concerned who proceeds to commit suicide also has a consciousness that the performance of the act is beneficial. The actor definitely has a consciousness, but no doubt an erroneous one, that the performance of the act of suicide is not a greater evil. Another important thing about the awareness of an act as beneficial is that it must be understood as relative to the time and circumstance. Here it has been conveyed that what is good in one circumstance at one time may not be so for an agent in another time and circumstance. What I desire to perform in one situation I may desire to avoid in another situation in life.

The third condition as mentioned by the Naiyāyikas in accounting for volition is a very complicated and difficult notion. The Naiyāyikas themselves are divided into two camps in interpreting this complex condition. According to one interpretation as offered by some leading Naiyāyikas, it is the consciousness of the absence of evil or undesirable consequences. On the other hand, according to others including Visvanath, the author of *Siddhantamuktavali,* this condition is a negative one as it implies the absence of consciousness about an act (towards which the agent is motivated) being productive of evil consequences. In defense of negative interpretation, Visvanath

brings in certain empirical considerations in favour of it. He cites instances of actual experiences when an actor has volition in spite of his not having this sort of positive consciousness of the absence of evil. However, Visvanath has nothing that tends to deny that in other cases there may be positive consciousness of the absence of evil immediately preceding volition as a condition. The Naiyāyikas, who do not favour negative interpretation, maintain that to admit the absence of consciousness of evil as a condition is to make it psychologically rather unintelligible as a condition of willing. The logical aspect of this debate can be clarified by borrowing the notion of the scope ambiguity of logical operators as found in Russell's *Principia Mathematica.* A sort of ambiguity is observable in this case between an epistemic concept and the concept of negation in respect of each of its scope over the other. The negative interpretation favours the dominance of the concept of negation over that of the epistemic concept; whereas, in the latter interpretation of the two concepts, the epistemic one predominates over the negative concept. This is not to be regarded as a mere terminological or stylistic variation for it has some very significant implications.

The Naiyāyikas did not stop pondering about that which actually deters or counteracts a volitional process from switching on. A thorough discussion is required of the nature of the deterrent that counteracts the occurrence of an effect in spite of all causal conditions being there. Two alternative explanations are possible in this case: according to one view, the consciousness of evil consequences that acts counter-productively, is the subjective feeling of aversion of the agent; in another view, it may be said that there is some object which is of evil nature that inspires the subjective feeling of aversion. Of these two possible interpretations, the Naiyāyikas favour the first one, according to which it is the subjective feeling of aversion that acts as the deterrent. Thus for the Naiyāyikas it is the subjective state of the agent that is important. The feeling that deters volition is relative to the person, time and circumstances; hence it can never be held to be absolute. This can explain adequately the difference in our attitude regarding various acts. I may be strongly inclined to perform an act Y, whereas Mr. A may forcefully refrain from performing the same act. Again it may also happen that it is the same I that may perform one act in one circumstance but not in another set of circumstances. The Nyāya account of volition thus presupposes as its condition certain complex entities comprising cognitive, conative

and affective aspects.

A more detailed analysis of the nature of the deterrent is available in the writings of the later Naiyāyikas. At least five alternative accounts about the nature of the deterrent are possible. Two very simple accounts are as follows: Firstly, it may be taken plainly that the cognition of the absence of any evil consequence is sufficient and along with the cognition of the action as producing good consequences plus resulting desire etc is the cause of volition. In this account more emphasis is laid on the intellectual side of the mind and on the positive account of the nature of volition.

Secondly, there is the possibility that that which deters volition is cognition. An agent may have some positive awareness of an act as leading to harmful consequences. Hence this consciousness of an act as producing evil consequences deters the agent from the performance of that act. In this case the absence of consciousness of an act as producing evil consequences may motivate an agent to will to perform that act.

These two very straightforward accounts of the nature of human volition may be replaced or supplemented by an account that projects the deterrent as a compound made up of two components, viz., one of feeling of aversion and the other is the cognition of evil consequences. In this view, neither mere cognition of evil consequences nor mere feeling of aversion is effective in deterring volition. If there is only a feeling of aversion but no cognition of evil consequences with regard to the performance of an act then also volition may occur. This third view considers the deterrent to be a compound condition consisting of both cognition of a consequence as evil and also the feeling of aversion towards such an act. This sort of deterrent is very effective in arresting volition. This may be considered as the necessary condition for the non-occurrence of volition. In the earlier mentioned two cases the occurrence of the volition and for that matter its non-occurrence is a possibility for it may occur or may not occur.

There is also another fourth alternative account of the nature of the deterrent. Like the preceding one this alternative also considers both the feeling of aversion and the cognition of evil consequences as the deterrent. Unlike the third alternative, here it has not been assumed that it is a compound of these two that deters. This view gives priority to the cognition of evil consequence with regard to an act in its origin. This primary epistemic state of mind in its turn gives rise to the feeling of aversion towards the act which is cognised to produce evil consequences. In this account a causal relationship is assumed to exist between these two different aspects of mind which together thwart volition. The cognition of evil consequences alone without being followed by the feeling of aversion is ineffective in counteracting volition. Again the mere feeling of aversion unless preceded by the cognition of evil consequences remains ineffective.

In the opinion of Dinakara, one of the ablest exponents of the Nyāya position, the cognition of evil consequences is the primary factor in causally determining volition. But that does not mean that the deterrent is the mere cognition of evil consequences. In some cases the cognition itself is sufficient, again in some other cases the mere cognition of evil is not sufficient. This may vary according to circumstances. But as primacy is accorded to the cognition of evil consequences the negative condition of volition is always the cognition of the absence of evil consequences and not only the absence of the feeling of aversion. This view thus denies the interpretation of the third condition as given by Visvanath. This view categorically asserts that the cognition of the absence of evil consequences has to be present to cause volition for merely the absence of a feeling of aversion is not sufficient. In working out all these alternatives the Naiyāyikas appeal to the perspective of a person who is supposed to report what passes in his mind when he has volition towards an act.

The Naiyāyikas do not recognise any casual relationship between the conciousness of good regarding an act and the cognition that it is to be done. The relation between the consciousness of good and the cognition that is to be done is a relation of the sameness of the object. In other words, the object which is the object of consciousness of good, is the same object as the object of consciousness of duty. The Naiyāyikas here do not like to postulate two psychoses, one conditioning the other, but only one psychic compound with two aspects, one of conciousness of good and the other, cognition of duty with reference to it. The Naiyāyikas do not admit, to put it more explicitly, any Niyamaka or the conditioning relation here between the two psychoses. From the above discussion it also clearly emerges that nothing is good or beneficial, except in relation to a subject. In this analysis proper significance is attached to the subject of volition

that determines his own values. This account in no way lends any support to mechanical determination. The Naiyāyikas give what may be called a self-concept-based account of motivation. The Prabhākara Mimāmsakas also agree with the Naiyāyikas in this respect in giving an account of motivation in this direction. In their writings an elaborate account of the psychology of human volition is given. But there is one major point of difference between the Naiyāyikas and the Prabhākara Mimāmsakas in the way they develop their total theory. The Prabhākaras, unlike the Naiyāyikas, are not consequentialists and they do not consider the consciousness that the act is conductive to a good or beneficial result as one of the conditions for volition. Rejecting the idea that the real cause for volition is anything external, the Prabhākaras rather project self itself as identified with or specified by some act to be done as causing volition. Thus, according to them the actor, before he is actually motivated to act, must have knowledge of himself as one who is qualified by the capability to perform that particular act. Criticizing the Nyāya approach the Prabhākaras claim their theory of inducement to action is superior to the Nyāya model and is more acceptable than it. One of the merits of this view is that as it shuns the consequentialist outlook it can account for all instances of motive for action whether it has some end or consequence or not. This is particularly interesting with regard to such acts which do not lead to any end. In Indian philosophy the possibility and nature of a particular act which does not have any end or result is a matter of great debate. However the Prabhākaras also do not deny that empirical acts lead to certain ends or consequences and in these cases some belief about the end as beneficial is there, but that is not the real motive that moves us. In empirical cases also this desire for attaining an end moves the will not independently, but as specifying the self or as appropriated by it. Hence they maintain that in all cases of volition we have two distinct psychoses, one leading to the other, and thus having a kind of determining relation between the two. The Prabhākaras regard this relation as that of the ground (premise) and the grounded (conclusion).

Indeed there is a long history of debate between the Naiyāyikas and the Prabhākaras on this issue which is being left out of this discussion. The Naiyāyikas and the Prabhākaras, however, in spite of their differences in detail, agree in spirit in projecting a psychological account of human volition. This account of volition in terms of the

psychology of the human mind has some significant metaphysical presuppositions. In Gotama's *Nyāya-Sūtra* it has been categorically asserted that desire, aversion, volition, pleasure, pain and cognition are the marks of self. Thus desire which is one of the factors conditioning volition is accepted as the characteristic feature of the self and nothing else can have it. Further the Naiyāyikas argue that desire etc. also prove the existence of an immaterial substance. This view goes against the views of the materialist Carvakas. But this view of self also goes against the views of the Buddhists who believe in the doctrine of universal flux. Buddha, however, had made the statement that the willed or volitional act alone is ethically significant. Buddhists also maintain that will is the distinctive endowment of humans by virtue of which they are moral beings. The Naiyāyikas urge that the Buddhist doctrine of universal flux leaves no room for voluntary activity. The main argument here is that since all voluntary activity is inspired by some conscious end which the agent seeks to attain, the admission of the doctrine of universal flux fails to account for any active endeavour of a conscious being the result of which will not accrue to himself but to another person. Hence unlike the Buddhists, the Naiyāyikas urge that there is some definite error in the supposition that the belief in the existence of a psychical continuum is a construction or the result of abstraction. Rather the Naiyāyikas enforce the existence of a permanent self as a distinct category. This is, however, just a very brief resumé of a very long story.

The Nyāya account of volition surely adds a feather to the cap of him who wants to establish consciousness as something unique. My objective here is to make one aware that there is a possibility of accounting for consciousness without holding that mental conduct words are definable in terms of concepts referring to physical/ chemical composition. This view clearly suggests that to know myself as an agent who has a particular belief, desire, preference etc involves knowing something quite different and also more complicated than the physical / chemical composition of a machine. The admission of self or soul is often viewed with utter disregard and great suspicion because of certain of its ecclesiastical implications. The difficulty in convincing one of the existence of a substantive soul should not necessarily make us overlook the possibility of the non-bodily and the non-reductive aspect of a person.

REFERENCES

Bhattacharya, Visvanath Nyāyapancanana. *Karikavali with the commentaries Muktavali, Dinakari, Ramrudri*. Edited by Atmaram Narayan Jere. Varanasi: Krishnadas Academy, 1982.

Maitra, Susil Kumar. *The Ethics of the Hindus*. Kolkata: University of Calcutta, 1963.

Perspectives from Western Philosophy

What is Hard about the Hard Problem?

WILLIAM JOHNSTON

Much of the recent work in the philosophy of mind in the West has focussed specifically on the relationship between mind and consciousness. One of the seminal contributions to the debate was the paper by Thomas Nagel, "What is it like to be a bat?" back in 1974. Since then there have been a number of philosophers who have worked on the ideas that were sketched out by Nagel – including of course Nagel himself. The basic claim made in that paper was that whilst progress can undoubtably be made in explicating and clarifying many of the features of the mind-body relationship, and indeed such progress has undoubtably been made recently in the neurosciences and cognitive sciences, there is one area where progress seems to be stymied. This is the attempt to explain the nature of consciousness itself. As Nagel commented at the beginning of "What is it like to be a bat?", "Consciousness is what makes the mind-body problem really intractable. Perhaps that is why current discussions of the problem give it little attention or get it obviously wrong".

It is certainly now untrue that the problem gets little attention, and there are many who would dispute the claim that their solutions to the problem are 'obviously wrong', yet the central claim made by Nagel as to the intractability of the problem, whilst it has exercised philosophers, does not seem to have received the attention that it deserves from others working in the field. In what follows I attempt to spell out some of the ways in which philosophers have attempted to extend and develop Nagel's basic idea that consciousness is not tractable neuroscientifically. In particular I examine those elements of the problem that have been seen as contributing to its intractability and make some suggestions as to the ways in which progress might lie.

Nagel asserts that, "To imagine a mental state from the inside would be what I have called an act of *sympathetic* imagination – putting

myself in a conscious state *resembling* the thing imagined – and it would be impossible to do this without putting my brain in a corresponding physical state" and, classically, therefore to imagine what it is like to be a bat would require that I put myself in the conscious state of a bat, which in turn requires adopting the appropriate physical state of a bat's brain, which since this is impossible for a human being (without of necessity *becoming* a bat), entails that I am unable to achieve the required act of imagination. I cannot imagine what it is like to be a bat, whilst at the same time being convinced that undoubtably there is something that it *is* like. Bat consciousness is real but is, for human beings, ineffable. "My realism about the subjective domain in all its forms implies a belief in the existence of facts beyond the reach of human concepts."

But Nagel has a more general point to make, namely that there is another and more significant imaginative leap that we are unable to make here, namely that of imagining how it is that the conscious state and the corresponding physical state are connected – not just for bats, but in all cases. It is this that makes the notion of consciousness intractable for Nagel. "If we acknowledge that a physical theory of mind must account for the subjective character of experience, we must admit that no presently available conception gives us a clue how this could be done. The problem is unique. If mental processes are indeed physical processes, then there is something it is like, intrinsically, to undergo certain physical processes. What it is for such a thing to be the case remains a mystery."

His basic notion that consciousness is not tractable because of our inability to come up with any intelligible conception of the mind-body relationship is nevertheless not one that carries with it any undue pessimism. This contrasts starkly with positions such as that held by Colin McGinn where the impossibility of resolving the intractability arises from the in-built limitations of the human mind itself – resulting in 'conceptual closure' for any explanation of the nature of consciousness for human beings. Nagel's position on the other hand is that whilst it is impossible to conceive how consciousness relates to the activities of the brain, it clearly does do so, and indeed one day we might know how. ". . . I believe it is not irrational to hope that some day, long after we are all dead, people will be able to observe the operation of the brain and say, with true understanding, 'That's what the experience of tasting chocolate looks like from the outside.'"

This then is what Chalmers has termed 'the Hard problem'. "Consciousness poses the most baffling problems in the science of the mind. There is nothing that we know more intimately than conscious experience, but there is nothing that is harder to explain." So what is it that is hard about the Hard problem? For Chalmers it is simply that, "It is widely agreed that experience arises from a physical basis, but we have no good explanation of why and how it so arises". The problem is one of conceiving what an adequate explanation of experience would even look like, let alone providing one.

There would appear to have been considerable misunderstanding of what is being claimed here. Pat Churchland for example argues that this formulation, "When critically examined . . . reveals itself to be unconvincing and counter-productive". One of her reasons for making this claim arises from her dissatisfaction with the 'I cannot imagine . . .' argument itself. "Whether we can or cannot imagine a phenomenon being explained in a certain way is a psychological fact about us, not an objective fact about the nature of the phenomenon itself. To repeat, it is an epistemological fact — about what, given our current knowledge, we can and cannot understand. It is not a metaphysical fact about the nature of the reality of the universe". This is not perhaps entirely fair on Chalmers, since his formulation of the problem is not couched in terms of conceivability, but is rather conceived as an attempt to make a distinction between what he calls the Hard problem and the Easy problems.

"When not much is known about a domain of phenomena, our inability to imagine a mechanism is a rather uninteresting psychological fact about us, not an interesting metaphysical fact about the world." Whilst McGinn would probably want to argue that this is actually a somewhat *interesting* psychological fact about us, a more serious point about her objection is that Churchland has rather oversimplified what is being claimed here. The relation between conceivability and possibility is more subtle than she allows. In particular we need to get somewhat clearer as to what we intend by 'conceivable', since whilst there are some senses of the term which clearly rule out any entailment from conceivability to possibility – "I can imagine discovering a triangle with four sides", or as in the classic example from Flew on the question as to whether it makes sense to conceive of being able to imagine being able to be a witness at one's own funeral. Flew's own robust answer is that this is not possible, since if one is witnessing a funeral, then you are alive, in which case

it can't be your own funeral. If it is your own funeral, then you are dead, in which case you can do no witnessing. But clearly one could argue that nevertheless such a scenario is imaginable. At this point things can easily descend into a pantomime of, 'Oh no you can't', 'Oh yes I can'. But there are other cases where what is imaginable is not so clear cut, and the relation between conceivability and possibility is far more unclear. Kripkean cases such as imagining a world in which water is not composed of H_2O is one such. But whether or not Churchland is correct in her assessment of the 'I cannot imagine' strategy, she is undoubtably right in her assessment that it is a crucial first step to accurately conceptualise the problem so we can ask the right questions and design appropriate experiments as a means to discovering a satisfactory solution. But this in itself, whilst good advice, does not dispose of the problem nor make it any more tractable.

But Churchland's main contribution to the debate here has been in her claim that the conceptualisation of mind-body problems into 'Hard' and 'Easy' is counter-productive. Whilst not denying that there is a genuine difficulty in understanding consciousness, she is reluctant to allow that consciousness is the only, or even main, area of difficulty here. ". . . difficulty *per se* does not distinguish it from oodles of other neuroscientific problems". She correctly points out that many of the so-called 'Easy' problems are in fact notoriously difficult and there are plenty areas, such as our understanding of information retrieval, where we lack key conceptual/theoretical ideas about how the nervous system performs, or indeed could perform, such functions. Not only that, but, as she is keen to remind us, an argument from ignorance is always fallacious. Given that we do not know something, the only fact that follows from this, is that we don't know it. Not that there is something unknowable out there. As she says, just because Aristotle could not imagine how a complex organism could come from a fertilised egg, this in no way undermines the possibility, or indeed actuality, of embryogenesis. All it does is tell us something about Aristotle and the world in which he lived.

But all this misses the point. Maybe the 'Hard'/'Easy' formulation was a mistake. Indeed in retrospect I suspect that many would agree with this, including perhaps Chalmers himself. Nevertheless it marks off something both real and important. Within the broad area of inquiry identified as 'the mind-body problem' there are clearly two radically different types of problems. There are, on the one hand,

those that relate to our understanding of various mental functions – such as information retrieval. Here, whilst there are real difficulties that we face in understanding the mechanisms by which this is achieved, nevertheless it is clear that once we have explained those mechanisms we have explained the function. As with all functions, from bottle opening to door closing, once we have explained the mechanism by which it is achieved, the only thing left to be explained is 'why?'. Why do we want to open the bottle or close the door? And why do we want to retrieve information? And these questions are not functional questions. They cannot be explained by revealing the mechanisms behind them. Why is the opening of the bottle often accompanied by a desire to do so? Indeed, why is any mental function associated with conscious experience? This is precisely the Hard problem. Calling it 'the Hard problem' is only misleading if by doing so we miss the key difference between Hard and Easy – which is not *per se* in the hardness or ease of answering them, but in the radical disparity between the two kinds of problems. In the one case – which applies in the overwhelming number of instances where explanation is called for – phenomena are explained by identifying function and structure. Whether this is bottle opening or information retrieval – these count as 'Easy'. In the other case there is the phenomena of consciousness – the desire to open the bottle or the wish to retrieve information – for which functional explanations are simply inappropriate.

Dennett has argued against this. He urges us to see that even in the case of consciousness, once we have explained the functions we have explained everything. There is nothing left over. Against this it has been argued that this may well be how things appear from a third person perspective, but the Hard problem is uniquely a problem that arises from and in the first person perspective. But more importantly there is a mismatch here between the claim that consciousness is not completely explained by explaining functions and Dennett's response that explaining functions explains everything. "Subtract the functions and nothing is left." A wine bottle has properties of mass, shape and colour. If one were somehow, *per impossible*, to remove the property of mass from the wine bottle, it is pretty clear that there would be nothing of any great interest left. But this certainly doesn't entail that there is nothing to wine bottles other than their mass. It is doubtless the case that if one subtracts all the functions associated with consciousness, one would be left

without consciousness. But this is no argument that therefore consciousness is function. Enjoyment of wine is predicated on the existence of wine. But this doesn't lead us to assume that wine and its enjoyment are identical. Sure, if wine did not exist we could not enjoy it, but there is more to the explanation of the enjoyment of wine, than the explanation of the properties of the wine. Experience gets left out of all functional explanations. We should not be surprised or hurt by this. We are happy to accept that there is a range of types of explanation available. Subjective experience needs to be explained, and a scientific reductionist account leaves everything of importance still to be explained. It is doubtless true that there is no more to Leonardo's 'Last Supper' than paint marks on canvas. But saying that Leonardo's 'Last Supper' is nothing other than paint marks on canvas, whilst true, is to leave everything that is of importance to us about it, unsaid.

So if we are not convinced by a purely reductive analysis of consciousness, the Hard problem remains. Why are neural processes accompanied by consciousness? And it will not do simply to assert that this is just how things are. That consciousness and neural processes are identical. As Chalmers has spelled out in some detail this identity statement has a very strange status indeed. One does not need the prior identity between corks and wine bottle stoppers to derive the fact that wine bottle stoppers are (often) corks. All that is needed is a functional definition of 'cork'. But in the case of the claimed identity between consciousness and neural processes this kind of a priori meaning is being posited rather than explained. Any identity claim requires an explanation of how it is that two apparently dissimilar phenomena are in fact identical. Without at least the possibility of explanation, the identity claim is empty.

So what exactly is it that makes the Hard problem so hard? There have been several recent attempts at answering this of which one of the most interesting is that of McGinn. He locates the difficulty in the non-spatial aspect of experience. Clearly we are going to have problems giving any kind of explanation in terms of the physical structure of the brain to phenomena that lack any spatial dimensions. McGinn is careful to note that this aspect of experience does not entail some non-spatial entity that *has* the experience, but even so it is difficult to see why the non-spatial aspect of experience is what makes its explanation so hard. Certainly it explains why it is impossible in any scientifically reductive mode, but we have plenty

of other examples of non-spatial properties, where the explanation is not especially hard. The effectiveness of the design of a wine bottle label is a pretty good non-spatial property, and yet we can give a good account of it terms of complex sociological and psychological factors. It seems to me that the non-spatiality of experience is certainly a key aspect of it that precludes certain kinds of explanations, but that in itself it does not pose any particularly distinctive problem. The problem lies elsewhere. It lies, it seems to me, in the difficulty of formulating the subjective quality of experience in a language. We simply do not have, as yet, a language adapted or developed to enable us to give formal expression to subjective experience – an 'objective phenomenology', as Nagel described it. The language in which we currently talk about experience is almost entirely derivative from the language we have for describing the external world.

If we were to develop a formalism through which we could describe, at least in part, the subjective character of experience in a form comprehensible to beings incapable of having those experiences, as Nagel suggests, this should, more importantly begin to reveal a phenomenology that may permit questions about the physical basis of experience to assume a more intelligible form. At the present time our talk about, for example, the bouquet of a particular wine, is arrived at from a first person perspective where the ease by which it is achieved and the lack of distance between experience and experiencer actually gets in the way of any deeper understanding of what is going on. One might try, for example, following the lead of Nagel, to develop concepts that could be used to explain to someone who lacked all sense of taste, what it was like to taste a fine wine. Clearly there is a sense in which we could not go all the way here. It is deeply unlikely that we could ever arrive at a situation where we achieve complete understanding, "Now I know what Nuits St. George tastes like!" But it ought to be possible to move in this direction. Maybe, again following Nagel, we could start with some of the structural properties of experience – particularly perhaps visual experience, where the structural components of the visual field ought to be amenable to formal expression. Whether or not this is a good place to start, it is clear that until we can actually express what the Hard problem is in terms that open it to access by the methods of science, we are going to remain in the present highly unsatisfactory situation where we either deny that the problem exists or alternatively that it is too Hard to be solved.

Is It Irrational to Talk to the Chinese Room?

STEFFEN BORGE

Introduction

The creatures of philosophy are a curious lot. There are among others swamp-man, twin-earth siblings, blockheads, brains in vats, Tibbles the-one-of-many-cats-cat, mindless duplicate zombies, otherworldly counterparts and the target of this paper: the Chinese Room. The Chinese Room example was originally introduced by John Searle (Searle, 1980) as a counterexample to the Turing test (Turing, 1950). All being a mind-haver amounts to, Turing suggested, is having decent conversational skills. The Chinese Room, as designed by Searle, had all that, but was by hypothesis not a mind-haver. The lesson to be learnt from the Chinese Room Argument, according to Searle, is that a system can be behaviourally and functionally indistinguishable from the real thing without being the real McCoy. The main thesis of strong Artificial Intelligence (AI) is that behavioural and functional features of a system suffice for having mental states, and that thesis is, according to Searle, false. This conclusion, however, has been controversial and there is presently among philosophers an intuition-divide regarding the status of the Chinese Room. This AI debate has developed into a trench war. In this paper I will present a diagnosis of the current stalemate of intuitions and suggest a new way of approaching the problem.

The participants in the debate have generally followed Turing and Searle in thinking that the fundamental question pertaining to the Turing test, and hence the Chinese Room Argument, is whether or not the system in question has a mind. This, I believe, is not the appropriate question to ask when approaching an artificial system like the Chinese Room. Rather the right question to ask is whether it is irrational to continue a conversation with a system like the Chinese Room after we have learned about its inner workings. When it comes to the question of mind ascription, this shift from metaphysics to

epistemology can, I suggest, provide us with a better understanding of how we can and should deal with possible upcoming successes of AI research. In the end my account is, I believe, a vindication of Turing; not in letter, but in spirit.

Turing's Imitation Game

In the opening of his paper, "Computing Machinery and Intelligence," Alan Turing wrote that he was to consider the question "Can machines think?". He then went on to remark that this question is not precise enough since the definitions of the terms "machine" and "think" might just reflect the current usage and thus suggest that the question should be settled by a Gallup poll. "Absurd," Turing exclaimed. Later on in the paper he went on to deem the original question not only meaningless, but "too meaningless to deserve discussion". In its place Turing suggested an operational test – the imitation game – as a guide to whether a machine should be attributed with a mental life.

Turing envisaged a game played by three people, a man (A), a woman (B) and an interrogator (C) communicating with each other via a teleprinter. The goal of the game is for the interrogator C to distinguish the woman given that they both (A and B) present themselves as one. There are no limitations to what sort of questions the interrogator can ask. Turing went on to ask:

> We now ask the question, 'What will happen when a machine takes the part of A in this game?' Will the interrogator decide wrongly as often when the game is played like this as he does when the game is played between a man and a woman? These questions replace our original, "Can machines think?" (Turing 1950: 434).

The Turing test purports to draw a line between accidental physical features of thinking things like us, and our intellectual capacities, and Turing's suggestion is that if a machine on a regular basis manages to trick the interrogator then it should be regarded as a thinking thing, the point being that it would be unreasonable for us to doubt the intelligence of a system that could pass such a complex and demanding test as the imitation game. We are after all allowing the interrogator to ask questions of any nature, so a machine that passes such a test must be substantially more complex than specialised machines that could only pass a very limited Turing test – like for

example Deep Blue. The basic assumption behind the imitation game is – as Dennett puts it, Nothing could possibly pass the Turing test by winning the imitation game without being able to perform indefinitely many other clearly intelligent actions (Dennett: 1998, 6).

A few remarks are needed. First, we might want to moderate Dennett's claim by stressing that the question of whether something could possibly pass the Turing test without being intelligent is – as we shall see – a highly metaphysically charged claim. What Dennett should claim is that it would be unreasonable to doubt the intelligence of such a machine. Second, failing the Turing test does not establish that the machine is not intelligent. The emphasis is on the conditional, "If it passes the Turing test, then it is reasonable to assume that it is intelligent". Third, and most importantly, talk of imitation is a red herring – one that Turing himself is responsible for. The word "imitation" is too easily conflated with or understood as "simulation" and here we can let the thousand counterexamples bloom. We have for example no inclination to regard a computer simulation of a tornado as being on par with a real tornado, so why should a computer simulation of thinking be any different? In both cases, one might argue, the machine is only abstractly going through the motions.

This point is well taken, but is due to an unfortunate choice of words. The game should not be viewed as a case of imitation, but rather as a question of competent performance. A computer simulation of a tornado does not perform as a real tornado, but if it did, in that it created chaos and mayhem over vast areas, then we would indeed call it a tornado. And the same goes for being a thinker. The machine if it is successful is not imitating a conversation but engaging in one – its conversational contributions have all the right effects and consequences.

This might be all well and fine, but why settle for the less ambitious question of what it is reasonable to count as having a mind instead of going for the big fish of answering the question of what it is to have a mind *simpliciter*? There are a few theoretical advantages. (1) It gives us an edge when it comes to justified mind ascription since the question of what it is to have a mind *simpliciter* is perhaps intractable. It could be that the mind-matter connection is in principle unsolvable – due to certain metaphysical considerations. Perhaps Colin McGinn is right when he argues that human beings are cognitively closed when it comes to understanding the psycho-physical link between a mind and the material that embodies it (McGinn, 1989). Just as penguins

are not equipped with minds or cognitive abilities to understand Gödel's incompleteness proof even though other creatures (like us) can, our cognitive constitution, according to McGinn, prevents us from understanding the mind-body interface. We might in a less pessimistic mode say that there is at least presently no widely accepted theory of how to discover whether a system other than yourself has a mind without consulting its behaviour and/or functional make-up. The Turing test – at least so far – is the only game in town.

(2) If we are interested only in the human species as our research topic, then we might do well in taking as a working hypothesis that human mental states are identical with certain brain states, but this will not do if we are interested in mental states in general. Why should we suppose that only systems that are made of the same chemical material as us could possibly have a mental life? Such a line of argument would be nothing but species-chauvinism displaying a hopeless anthropocentric attitude to the phenomenon of thinking. Minds, one might argue, are conceivably multiply realisable, and then what matters is not what a system is made of, but how it functions or behaves.[1] We might furthermore believe that mental states could be not only realised in multiple materials, but also that a variety of different functional organisations could give rise to intelligence. The Turing test is open to these possibilities. No undue bias is displayed against specific species or material constitutions, or – and this is where the Turing test is more liberal than many functionalistic theories of the mind – particular functional organisations.

Searle's Chinese Room

Only game in town or not, the Turing test – and with it strong AI – has met resistance. Among its harshest critics is John Searle. By taking his starting point in the performative essence of mind-having, as advocated by Turing and his followers, Searle tries to show how this feature cannot take centre stage when it comes to the question of having a mind. Searle invites us to consider the following thought-experiment (Searle 1980: 417-418). Searle, a non-Chinese speaker, is locked in a room together with an instruction manual in English that pairs Chinese symbols with other Chinese symbols. The book enables Searle – when he has internalised the instruction manual – to immediately respond to any incoming string of Chinese characters by sending out another string of Chinese characters. Unbeknownst to Searle the

incoming strings of symbols represent questions, while the outgoing strings of symbols represent answers to these questions. Searle's performance is, for the people outside the room, indistinguishable from that of a native Chinese speaker. Searle passes the Turing test for speaking Chinese. But can we from that conclude that Searle knows or understands Chinese? No. Searle is able to manipulate symbols in accordance with a set of instructions, a formal program, but he has no understanding of the activity he is engaged in. And without such understanding he cannot be said to speak or understand Chinese.

By parity of reasoning the same holds for the machines in Turing's imitation game. The proponents of strong AI that claim that having a mind is nothing but the implementation of a certain computational program – e.g. the right functional organisation – are wrong since the Chinese Room shows that such an implementation does not suffice for Searle to know Chinese nor would it suffice for a computer working with a formal program to have understanding. Searle in the room comes over as a competent Chinese speaker and we could elaborate the example so that the inner workings of the Chinese room – the way Searle provides the output – is functionally iso-morphic to that of an ordinary Chinese speaker.[2] Searle only manipulates what are for him meaningless symbols, and in the same manner a computer is merely engaged in mechanical operations. Any system that is merely following a rule-governed procedure – an algorithm or formal program in regard to its inputs and outputs – fails to have intentionality and consciousness. Searle in the room has no semantic understanding of the activity he is engaged in. Syntax does not suffice for semantics and since computers, as defined by Searle, merely instantiate a formal program, they have syntax, but no semantics. The argument is summed up in the following way (Searle 1994: 39):

> Premise 1. Syntax is not sufficient for semantics.
> Premise 2. Computer programs are entirely defined by their formal, or syntactical, structure.
> Premise 3. Minds have mental contents; specifically, they have semantic contents.
> Conclusion: No computer program by itself is sufficient to give a system a mind.
>
> Programs, in short, are not minds, and they are not by themselves sufficient for having minds.

What it is to have a mind is not captured by mere extrinsic features like functional input-output relations or performance. Such machines would lack what Searle calls *intrinsic intentionality* and thus have no minds.

But what is intrinsic intentionality and how can we tell when someone has it? On the latter question Searle is rather quiet, but he has things to say about the former. At one point he merely stipulates that intrinsic intentionality is that something extra that makes native Chinese speakers have a semantic understanding of what they do as opposed to Searle in the room. It is what makes brains think while mere computer simulations of our neural brain structure fail to do so.

> So I am just stipulating that by "intrinsic intentionality" I mean the real thing as opposed to the mere appearances of the thing (*as-if*), and opposed to derived forms of intentionality such as sentences, pictures, etc. (Searle 1995a: 80).

There is no need to fault Searle for just stipulating this. Even though the notion of "intrinsic intentionality" is rather vague, we all have a certain intuitive grasp of the distinction he is making. I – and hopefully you – have intrinsic intentionality, but when I describe my car as wanting more gasoline I speak merely figuratively. My car has nothing but as-if intentionality and the fuel gauge that tells me how much gasoline there is in the tank can only be ascribed derived intentionality. The fuel gauge's intentionality derives from our usage of such devices to tell us something about the world. Elsewhere Searle gets somewhat more specific about intrinsic intentionality. Only things with the right causal powers have intrinsic intentionality and thus minds.

> My own view is that *only* a machine could think, and indeed only very special kinds of machines, namely brains and machines that had the same causal powers as brains (Searle 1980: 424).

> Another misunderstanding of the Chinese Room Argument is to suppose that I am arguing that as a matter of logic, as an a priori necessity, only brains can have consciousness and intentionality. But I make no such claim The point is that we know in fact that brains do it causally. And from this it follows as a logical consequence that any other system that does it causally, i.e. that produces consciousness and intentionality, must have causal powers to do it at least equal to those of human and animal brains

powers to do it at least equal to those of human and animal brains (Searle 1995b: 547).

Searle's position is not that no artificial or non-carbon based system could produce a mind. He is merely insisting that a syntactic machine that was nothing but a syntactic machine could not – due to the Chinese Room Argument. Whatever it would take for other systems than us to have minds, they would minimally have to have the same causal powers as the human brain. The keyword is "same causal powers" and not functional organisation or behavioural input-output relation. What proponents of strong AI ignore, according to the Searlean picture, is that the matter matters when it comes to the question of having a mind.

The stalemate of intuitions

The reaction to the Chinese Room Argument has been divided. The philosophers who share the Searlean intuition take the Chinese Room Argument to show that behavioural features and functional capacities could not establish a system as thinking. Both behaviourism and functionalism fail to provide sufficient criteria for mind ascription. Both theories fail – at the end of the day – to distinguish between systems with intrinsic intentionality from those without. The question of what if anything could function as an observable criterion for regarding systems as having intrinsic intentionality, however, is often left unanswered. One might claim in a rather unsympathetic manner that the intuition that it is obvious that the Chinese Room lacks some crucial feature has left the argument's followers with little else than Cartesian certainty about their own mind and a claim that it is reasonable to assume that other humans – creatures with the same wetware – also have a mind since they share our physiology. And that misses the point that the Turing test, and presumably the Chinese Room Argument, were about the ascription of mental capacities to systems with a physical constitution different from ours. The reasonable assumption that the human brain is responsible for us having a mind is not going to cut any ice when it comes to evaluating systems with different material constitutions.

On the other side of the intuition divide bullets were being bit. The Chinese Room Argument clearly shows that Searle in the room does not understand Chinese, but does it necessarily show us anything

more than that? Even though Searle in the room is not an eligible candidate for understanding Chinese, might there not be other candidates for having the required understanding? We could, for example, say that perhaps the room as a whole understands Chinese.[3] Who would initially when you look at it take the grey slimy matter we call a brain to be a plausible candidate for being the main causal contributor to intrinsic intentionality? Yet we know it is. Could not the same be the case with the Chinese Room? There is a crucial difference between a system as a whole and its parts. Most of us would either say that brains think or that things with brains think, but we would not say that individual neurones or smaller parts of the brain think. What matters is the system as a whole or larger parts of it. This then could account for why our intuitions immedi-ately side with Searle's Chinese Room Argument and why our considered view, in light of our intuitions about brains and their parts, should be that the whole thing could possibly understand Chinese or be a thinking thing. Searle in the room does not under-stand Chinese, but he is just a small part of the system. It could also be that the resistance we find towards entertaining the idea that systems like the Chinese Room could have mental capacities is just a hard-wired biological reflex. But such inbuilt biological prejudices should be given no philosophical weight.

If we furthermore ask about the locus of intrinsic intentionality we see that not even the brain is a well-suited candidate for being that. We would like to say that I can truly utter that "I have a hand", but that comes out as false if the brain is the locus of intrinsic intention-ality, since the brain has no hand. We can observe that when we press the question of what the locus of intrinsic intention-ality is, we find that most of the answers presented by philosophers are rather counterintuitive. This observation might mitigate some of the counterintuitiveness we feel towards the the-room-as-a-whole-thinks solution.

Unfortunately this line of argument does not provide us with any new criteria beyond behavioural or functional features for when it is reasonable to regard a system as thinking and when it is not. The argument does not go any further than what can already be read out of the initial Turing test and for that reason it has failed to move anyone on the other side of the intuition divide. This is where the debate has been stranded since Searle originally introduced the

Chinese Room Argument and this stalemate of intuitions gives us motivation for approaching the question from a different angle. Before we can do that it will be useful to look at a few different principles for mind ascription.

Three types of principles for mind ascription

Let me briefly sketch three different methodological principles for mind ascription. These principles – I believe – cover most of the conceptual terrain, even though they might not be exhaustive.

(i) *Mad-Dog Mind Ascription* (*MMA*): There is no fact of the matter pertaining to the question of whether a system has a mind or not. Demanding that there should be some discernible fact of the matter is – to echo Turing – all too meaningless. Whether or not something has a mind is a question of interpretative choice. That choice is presumably based on pragmatic concerns about winning the survival game or whatever other game one is playing. This means, in effect, that there is nothing in principle wrong about attributing mental states like desires and beliefs to, for example, a thermostat, if that helps us elucidate its behaviour. There are nothing but *as-if minds*, or, *any system can be said to have a mind* if we choose to view things that way. Dennett and McCartney have, at times, come close to this kind of view (Dennett, 1981, McCartney, 1979).

(ii) *Honest-to-God Mind Ascription* (*HMA*): There is a fact of the matter pertaining to the question of whether a system has a mind or not. This fact, whatever it turns out to be, is independent of functional and behavioural evidence. This question could furthermore, at least in principle, be decided in disregard of such evidence. The question of whether something has a mind or not is a matter of ontology – of what there is – and not a matter of epistemic principles that state what it is reasonable to believe. This is, among others, Searle's position.

Both these models share an all or nothing attitude towards the question of the importance of the notion of "intrinsic intentionality". Intuitively, the *MMA* model seems highly unsatisfactory. There must be more to having a mind than merely being viewed as such. We would like to be able to say that certain things do not count as thinking even if it might at times be advantageous to view them as thinking. The pragmatism of *MMA* is unsatisfactory since it cannot tell us why it is unreasonable to view some systems as thinking, only why nothing

necessitates such a judgement. The following story brings out this point.

> In lectures Wittgenstein imagined a tribe of people who had the idea that their slaves had no feelings, no souls – that they were automatons – despite the fact that the slaves had human bodies, behaved like their masters, even spoke the same language.
>
> Wittgenstein undertook to try to give sense to that idea. When a slave injured himself or fell ill or complained of pains, his master would try to heal him. The master would let him rest when he was fatigued, feed him when he was hungry and thirsty, and so on. Furthermore, the masters would apply to the slaves our usual distinctions between genuine complaints and malingering. So what could it mean to say that they had the idea that the slaves were automatons? Well, they would *look* at the slaves in a peculiar way. They would observe and comment on their movements *as if* they were machines. ('Notice how smoothly his limbs move.') They would discard them when they were worn and useless, like machines. If a slave received a mortal injury and twisted and screamed in agony, no master would avert his gaze in horror or prevent his children from observing the scene, any more than he would if the ceiling fell on a printing press. Here is a difference in 'attitude' that is not a matter of believing or expecting different facts (Malcolm 1954: 548-549).

If having a mind, as *MMA* seems to insist, is merely a question of viewing the system in a particular way, e.g., a willingness to view the creatures in question from an intentional viewpoint, then one can, of course, refuse to adopt such an attitude. The requirement of this willingness explains how it is possible for Wittgenstein's master-tribe to view their slaves as automatons. By not adopting a certain attitude towards the slaves, the fact that the slaves do have minds does not come into focus for the master tribe. But *MMA* fails to explain why Wittgenstein's master-tribe is wrong or mistaken to view the slaves as automatons.

HMA theorists do better here, since they will insist that having a mind is a question of having intrinsic intentionality and it does not matter what attitude the masters have towards their subjects. The masters are just wrong. One might, however, object that the *HMA*

approach to mind ascription only makes sense insofar as we can make sense of a God's point of view in regard to our position as interpreters of other minds. God might be able to say which functional organisations in what sort of materials can yield a mind and thus be able to distinguish between systems that are otherwise indistinguishable with respect to mental properties. But does this point of view make sense for creatures like us? *HMA* offers a point of view that seems, at least presently, inaccessible to us. And what reasons do we have for assuming that we will ever be able to deduce from a set of premises about a system's material constitution, perhaps together with functional features, whether or not the system has a mind? *HMA* seems, on the face of it, to be too strong, while *MMA* is obviously too weak. An intermediate position is called for.

(iii) *Benefit-of-Doubt Mind Ascription* (*BMA*): One is *prima facie* entitled to ascribe a mind to every system that comes over as thinking. Systems, no matter how different from us in form, shape and material constitution, should be given the benefit of doubt and be treated like thinking things if they have features like conversational skills that strike us as intelligent. In the master-slave story from Wittgenstein, we can say that whatever it is that makes the masters come over as thinking, the same sort of behavioural evidence is displayed by the slaves and that makes the masters unjustified in treating them differently. Notice that *BMA* is a default position and considerations about inner make-up might or might not override our initial entitlement.

BMA differs importantly from *MMA* and *HMA*. It differs from *MMA* in that it proposes restrictions on when we are entitled to ascribe mental states to a system – not everything goes. It differs from *HMA* in that it bypasses the direct ontological question of mind-having, and puts emphasis on our interaction with possible mind-havers and the entitlements that such interaction might transfer. But insofar as *BMA* bypasses the ontological question posed by *HMA*, then the two positions are compatible and it is not clear why I am pitting them against each other. The reason I contrast the two is not because they are incompatible, but rather because they disagree on the issue of the priority of what question one ought to ask when confronted with a system that comes over as intelligent. This connects with our initial question of whether it is irrational to keep on talking to a Chinese Room when you have learnt about its inner make-up.

Talking of the Chinese Room

Take the situation where you have logged yourself onto an internet chat-room and spent a long time talking to one particular participant. Imagine that it is then brought to your attention that you are talking to a computer – a Chinese Room. Do you continue the conversation? Note that the question is whether you continue the conversation *as before* and not whether you continue to respond in order to check out the sophistication of the computer software. If one opts for the latter option, then one has ended the conversation and is rather involved in some sort of reflex checking game. What then does it take to continue the conversation?

The most plausible answer to this question, I think, is that we need to have or display the right kind of attitude towards the system we are talking to. In the cases under consideration the system comes over as thinking, since it carries its part of the conversation. Following *BMA* we are *prima facie* entitled to regard it as a mind-haver and in doing so we adopt a *reactive attitude* towards it (Strawson, 1982). In adopting a reactive attitude towards a sign-producing system one forsakes an *objective attitude* where the system is seen as merely confer-ring information, in favour of seeing the conveyed information as a consequence of the system holding certain propositions as true. A reactive attitude furthermore makes us inclined to view the system as responsible for its conversational moves, and we are willing to judge its actions in normative terms. We expect our conversational partner to follow certain conversational rules and practices, and failure to do so is met by sanctions.

The entitlement to a reactive attitude that we, according to *BMA*, are justified in adopting is a default entitlement. The question now is whether our entitlement has been defeated in the Chinese Room Scenario when we learn about its inner workings? This question is importantly different from the question Searle asked us to consider (Searle 1994: 36):

> The question we wanted to ask is this: 'Can a digital computer, as defined, think?' That is to say: 'Is instantiating or implementing the right computer program with the right inputs and outputs, sufficient for, or constitutive of, thinking?' And to this question, unlike its predecessors, the answer is clearly 'no'.

We can rewrite this, in light of our present discussion, as follows:

The question we now want to ask is this: 'Can we retain the reactive attitude we have adopted towards a system when we learn that it is a digital computer?' That is to say: 'Is instantiating or implementing the right functional capacities with the appropriate conversational output to conversational input, sufficient for, or constitutive of, a justified reactive attitude towards the system?'

The answer, I think, is no longer obvious. People's inclination to say "no" is weaker in the latter case than when we tell John Searle's Chinese Room Story. If your intuitions still side with Searle, then let me present the following three scenarios. See if your reactions to the three cases differ. *Case (i)*: You just logged yourself onto an internet chat-room and someone writes something that you find offensive. You write your first entry:

You: That's not very nice.
(You are then made aware of the fact that your conversational partner is a computer.)
Chinese Room (*CR*): What do you mean?
You: Oh, never mind.
Log off.

Case (ii): You logged yourself onto an internet chat-room and spend hours conversing with one particular participant. Again it is brought to your attention that you are talking to a computer. The following conversation takes place:

You: Are you a human or just a machine?
CR: What do you mean by *just* a machine?
You: Well. Somebody built you and programmed you to do what you do now.
CR: Well yeah I guess I'm what you would call an artefact. For me, of course, it is quite natural since it's the only thing I know. And sure I'm programmed. There are limitations to my operations, but these give me plenty of elbowroom. I've heard that there are certain cognitive limitations to you guys too.
You: No, no, no. That's beside the point. I have thoughts, feelings, subjective attitudes, you can't have that; you're a machine.
CR: How do you know that?

At this point you are behaving as if the Chinese Room is a person, while charging it with not being one. Would you log off at this point in case (ii) in order to be coherent?

Case (iii): Again you have come in contact with something in an internet chat-room. This time, however, the relationship runs over years. You both regularly log on for chats, send each other emails, exchange Christmas and birthday cards, etc. Would you end what you for years have regarded as a valuable and intellectually stimulating friendship, in face of the information that your convers-ational partner is a digital computer?[4]

Even if your initial intuitions sided with Searle would you not, at least, admit that your intuitions in regard to case (iii) is no longer so clear? Does the idea of continuing the conversation *really* strike you as nonsensical or even irrational? Searle has a tendency of charging anyone that does not accept the purported conclusion of the Chinese Room Argument with being in the grip of a theory.[5] Now it seems that the tables are turned. If the new way of approaching the question and our three examples do not at all alter or touch your intuitions, then I fear that nothing will. I suspect that no arguments or examples whatsoever would make you consider any other systems than human beings as being possible candidates for having a mind.

If you, on the other hand, notice a change in your intuitions about the Chinese Room or attitude towards the question posed, then the following problem emerges. The only relevant difference between cases (i), (ii) and (iii) is the time of acquaintance. If you are tempted to apply a benefit of doubt principle about your reactive attitude towards the Chinese Room in case (iii), then that should carry over to case (i) and (ii). Perhaps the machines in case (i) and (ii) prove themselves to be too crude for you to retain your reactive attitude. That, however, does not mean that you should not adopt a reactive attitude even when you know about their inner workings just because you have only interacted with them for a short period of time. If you accept this then it seems like you should retain a reactive attitude toward the Chinese Room in Searle's original example. It does not seem irrational to talk to the Chinese Room. Why should you then deny that is has a mind?

One might reply that intuitions are not a good guide to mind ascription. Just because I cannot help myself but adopt a reactive attitude towards some formal systems, due to their cleverness and sophisticated conversational performance, it does not mean that I am forced to regard them as genuine thinkers. Indeed it does not. But then you leave us in the dark about what will, if anything.[6] One might, however, challenge the picture I have painted by arguing that I am

conflating metaphysical and epistemological concerns. Let us address this latter question.

From epistemology to metaphysics and back again

A critic might argue that it is not clear in the previous sections whether the mind-ascription theories are meant to be epistemic or metaphysical. That is, it is not clear whether the theories are theories of what it takes to have a mind, or of what it takes to make it reasonable for others to conclude that a system has a mind. If *HMA* is the first kind of theory, then it is not appropriate to question it for its epistemic uselessness. We all know that reasonable rules can lead to error some of the time. And as noted, *HMA* and *BMA* are consistent and it thus remains somewhat unclear why I choose to contrast them with each other.

This observation might be carried over to my debate with Searle and his Chinese Room Argument. Searle is clearly attacking the metaphysical thesis of strong AI: that satisfying some functional role is sufficient for thinking. I seem to be replying by defending a certain epistemic or pragmatic thesis: that it can be reasonable to believe of a machine that it is thinking, and that it can be worthwhile to have a "human-like" conversation with a machine. Do these claims of mine really pose a problem for Searle's position? The first clearly is no problem from Searle's perspective, because of the fact that reasonable epistemic rules can lead you into error, at least sometimes. Searle can admit this without giving up anything in the Chinese Room Argument. The second is more of a problem, but as it is formulated it is more like a challenge than an argument. This challenge could be met by simply saying that what makes us regard a system as a thinker when we learn about its inner workings is evidence of similar physiological make-up to our own. Searle though does not push this line of argument since he accepts the argument that mental states could be multiply realisable. The main loophole in my argument, the critic continues, is the question of how we get from my epistemic claim: it is reasonable to keep adopting a reactive attitude towards the Chinese Room even when we learn that the Chinese Room is a machine, to the denial of Searle's metaphysical claim that if the Chinese Room is a machine, then the Chinese Room does not think. A quick, but essentially solid way of countering this metaphysical claim is to say that if it is epistemically reasonable to believe that a system that passes the Turing

test has a mind, then it is reasonable to believe that the metaphysical thesis of the Chinese Room is false. But let us not be that quick with the Searlean position.

BMA is supposed to be an intermediate position between *MMA* and *HMA*, and is as such a blend of the two, capturing, hopefully, their best features. Searle is right, I believe, in pushing the point that for a system to have a mind it must have intrinsic intentionality – whatever that is. Or to put it differently: our concept of what it is to have a mind entails among other things that the system has intrinsic intentionality. Someone that, for example, is merely parroting words and sentences does not engage in a conversation, since speaking a language requires that one utters the words with certain communicative intentions, which again seems to require an understanding of the activity one is engaged in. If we knew that the implementation of a computer program did not cause intrinsic intentionality, then we would know that that machine did not think. But this is the bone of contention. How do we know that a specific implementation of a formal program in, say, a digital computer does not create a mind and thus something with intrinsic intentionality?

Recall Searle's argument that because computer programs are defined by their formal or syntactical structure they do not suffice as prerequisites for having minds – e.g. for having intrinsic intentionality. But we are not mainly interested in the program itself but rather the implementation of the program in some material. It does not follow by necessity from the fact that computer programs are defined syntactically that the implementation of them cannot suffice for semantics. David Chalmers pushes this point when he presents the following parody of Searle's Chinese Room Argument (Chalmers, p. 327):

> Premise 1. Recipes are syntactic.
> Premise 2. Syntax is not sufficient for crumbliness.
> Premise 3. Cakes are crumbly.
> Conclusion. Therefore, implementing a recipe is insufficient for a cake.

The conclusion, of course, is false. This parody brings into focus that we should, in this debate, concentrate on the implementation of a program and not only the program itself. For Chalmers' parody to work, however, it would have to concede the Searlean point that matter matters or that matter might matter. If a recipe for a cake was just

defined formally – that is, the recipe did not prescribe what materials one should use in implementing the recipe – then we could easily imagine an attempt to implement a recipe with materials that were not suited to create a cake.

Let us consider another example. An architect's design of a footbridge could be defined by its formal or syntactical structure – meaning that the architect has not specified the material with which the bridge should be built. If it is part of the definition of a footbridge and subsequently baked into the description of the bridge one has designed that at least one person can walk over it, then some materials can implement the design while others cannot. The test for whether a certain material can implement the footbridge design is operational – does the thing perform when put to the test or not? And, if some material *X* can implement a footbridge design, then any material with the same causal powers as *X* can. Hence, if both materials *X* and *Y* can implement the design then we know that they have the same causal powers to do so. The same goes for different kinds of minds. We know that human brains can implement a formal program so that the system has a mind. Our test for this is – with the exception of our own mind, which we are directly acquainted with – operational. Does the thing perform adequately when put to the test? What else could such an operational test be, but an elaborated Turing test? This is the feature that *BMA* adopts from *MMA*.

In all these cases – bridges, cakes and minds – there is an operational test that the system will have to pass before we say that a program has been successfully implemented in a specific material. In the two former cases, however, we know why the material can implement a certain formal program successfully – we know which microphysical facts explain why a certain material succeeds and why others fail. Furthermore we have no inclination in, say, the footbridge case, to say that some material could implement the design and thus pass the operational test, while still not counting as a footbridge. But perhaps this is where minds differ from things like footbridges?

We know that if a formal program is implemented in the right material for having a mind, then it has intrinsic intentionality. But what is the scope of "right material" for mind-having? We do not even know what particular features of our brains make this material adequate for implementing programs that yield minds. We have, furthermore, no argument for the impossibility of other causal

structures than those found in human brains producing minds. We know trivially that if some causal structure in some material is capable of producing a mind, then it has the same causal powers as our brains. Presently, however, the question of causal powers in regard to the question of having a mind only comes into focus by looking at the systems' performance. We judge that something most likely has the right causal powers because it performs as if it had such powers. This is mirrored in *BMA* by its emphasis on performance. The possibility that perhaps current day computers are not of the right material to support a mind – e.g. that the Chinese Room Scenario as envisioned by Searle might be a real possibility – is taken care of by making *BMA* a default position.

The emphasis on the implementation of a program takes the wind out of the Chinese Room Argument. Granted, the scenario described in the argument is indeed conceivable and thus perhaps logically possible, but that is not enough for Searle to make his case against strong AI.[7] That we can envisage such a scenario does not show that the actual implementation of the right formal program in silicon-chips of modern day computers would not constitute a mind-haver. Perhaps it would not, but, indeed, perhaps it would. If strong AI is taken to be a metaphysical thesis claiming that in no possible world, the implementation of a certain formal program that yields a system behaviourally (and functionally) indistinguishable from creatures like us could fail to have intrinsic intentionality, then the Chinese Room Argument retains its bite.

If, however, the thesis of strong AI is taken to be an empirical hypothesis about our world – as I think it should – then Searle's argument can no longer have the same force.[8] Perhaps a Chinese Room of the Searlean kind is nomically impossible – e.g. it is just an empirical fact about our world and the natural causal laws pertaining to it that we could not build a mindless Chinese Room. Perhaps this is a world where anything – or a wide variety of different material constitutions including that of computers - that implements a formal program of a certain complexity and which passes an appropriately sophisticated Turing test will inevitably have intrinsic intentionality.[9] The metaphysical question now turns out to be a question of what kind of world we are in – what our modal address is. It does not seem to me that any of the sides enjoys an upper hand. Consider the two parts of what I call the Modal Address Argument.

Modal Address Argument, part I

Premise 1. If I know that an actual implementation of the right kind of formal program in a computer – e.g. a program that enables the system to pass a sophisticated Turing test – also, inevitably, gives the system intrinsic intentionality, then I know that John Searle's Chinese Room Scenario is empirically (nomically) impossible.
Premise 2. I do not know that John Searle's Chinese Room Scenario is empirically (nomically) impossible.
Conclusion. I do not know that an actual implementation of the right kind of formal program in a computer – e.g. a program that enables the system to pass a sophisticated Turing test – also, inevitably, gives the system intrinsic intentionality.

Perhaps Searle is right when he claims that strong AI is misguided. The following argument, however, can be made in favour of the research program of strong AI.

Modal Address Argument, part II:

Premise 1. If I know that an actual implementation of the right kind of formal program in a computer – e.g. a program that enables the system to pass a sophisticated Turing test – could fail to give the system intrinsic intentionality, then I know that John Searle's Chinese Room Scenario is empirically (nomically) possible.
Premise 2. I do not know that John Searle's Chinese Room Scenario is empirically (nomically) possible.
Conclusion. I do not know that an actual implementation of the right kind of formal program in a computer – e.g. a program that enables the system to pass a sophisticated Turing test – could fail to give the system intrinsic intentionality.

The net effect of the Modal Address Argument is to show us that we do not know what our world-location (i.e. our modal address) is in regard to the empirical possibility of a Chinese Room Scenario. Since we are unaware of our world-location, strong AI is a perfectly legitimate research program to pursue. It also leaves us, so far, with *BMA* as our only guide for mind ascription and subsequently for giving an answer to whether computers in our world could think, and that is a vindication of Turing, if not in letter, then in spirit.

The relevance for AI research

What then is the relevance for serious researchers in the field of AI of these high-flying philosophical remarks? The upshot of the current discussion is that researchers in AI should not worry too much about arguments like the Chinese Room Argument. Thought-experiments like John Searle's Chinese Room help us to get clearer on how we understand the concept of a mind, and that is a legitimate and valuable philosophical concern. However, when the argument is purported to have empirical consequences for current AI research, the argument loses its validity. An argument along the Searlean lines to the effect that the implementation of the right formal program in, say, silicon chips in current day computers does not suffice for having intrinsic intentionality would have assume that our world is a world were there could be systems that were functionally and behaviourally indistinguishable from mind-havers like us, but which nevertheless lacked intrinsic intentionality. But that assumption begs the question against the empirical version of the strong AI thesis.

A priori considerations from the philosophical armchair will not tell us whether the more ambitious aims of the AI research program of one day creating a thinking computer is within reach or not. Furthermore it does not follow from the philosophical problems of understanding the nature of mind and how we could tell with certainty that a system has it, that the engineering problem of creating a mind is unsolvable or as hard as the philosophical problems have turned out to be. All we can do, I believe, is to do our science and place our bets to the best of our knowledge. And if I may allow myself to speculate, I believe that the question of whether computers can think will be answered the day – if ever – that we have a computer that passes a suitably sophisticated Turing test. Metaphysical considerations about the possibility of such a computer being nothing but a Chinese Room of the Searlean kind will be trumped, I conjecture, by the impressive sophistication of such a machine and our inability to come up with an argument that does not make us look too much like Wittgenstein's master tribe. Granted, a stubborn refusal to accept such a system as thinking can, unlike Wittgenstein's master tribe, point to the differences in physiology between the computer and us as an argument in its favour. But then we are back to square one and we are left with an unsupported species-chauvinism. If we resist such a move then I believe that we will find *BMA* to be the principle of mind-ascription by which we should live.

NOTES AND REFERENCES

1. The argument that mental states are multiply realisable was first presented by Putnam as a critique of type identity theories of mind (Putnam, 1960). See also Fodor, 1968.
2. Searle presses this point in Searle (1990: 28) in order to show that connectionist systems that mimic the cognitive processes of our brains also fall prey to the Chinese Room Argument.
3. This line of argument was already suggested or hinted at by several of the authors in the "Open Peer Commentary" that accompanied Searle's original 1980 paper.
4. We might, at this point, want to add that it is not the case that the machine during your interaction with it has mostly told you lies. That is to say that the machine has avoided certain topics that it would perhaps have been natural to discuss had your conversational partner been a human being.
5. See for example Searle (1990: 31): "It is best to see strong AI as one of the last gasps of this antiscientific tradition".
6. It might be, as I have mentioned a few times earlier, that we want to put limitations on what sort of functional organisations we would accept with regard to the question of whether something has a mind or not. According to this kind of argument the problem with the original Chinese Room example is that the way it produces its responses are to crude. This, of course, is not Searle's line (see footnote 2), but it seems be Ned Block's line in his "Psychologism and Behaviour". Block imagines a machine somewhat similar to ELIZA (see Weizenbaum, 1966), only a lot more complex. The Block machine's way of responding to conversational input is to search through a huge memory bank of stored conversations in order to find a conversation that matches the score of the conversation it is engaged in and then give the output respond it finds in that conversation (Block, 1980, pp. 19-24). Finding out that a machine that passes a suitably sophisticated Turing test is a Blockhead might override our initial *BMA* entitlement to conceive the Blockhead as a mind-haver. Dennett claims that Blockheads are "utterly impossible" (Dennett, 1998, p. 12. Again we should read Dennett's "impossible" as "empirically impossible"), since such programs exhibit the feature known as combinatorial explosion (see Newell and Simon, 1976). The question then is whether Blockheads are empirically possible or not, and I believe that what I have to say about the Chinese Room in respect to this question in section 7 applies equally well to Blockheads.
7. I write "perhaps logically possible" because there has in the slipstream of Chalmers' *The Conscious Mind* been a controversy about whether conceivability is a good guide to logical possibility, and whether we should distinguish not only nomic possibility from logical possibility, but also metaphysical possibility from logical possibility. There is no room in this paper to enter into this debate.
8. It is also reasonable to assume that strong AI is a thesis about what happens in our world and not about what happens in all possible worlds, especially when we remember that one of Searle's targets in his original 1980 paper was strong AI as a research program. See, for example: "The project is to reproduce and explain the mental by designing programs, but unless the mind is not only conceptually but empirically independent of the brain you couldn't carry out the project" (Searle 1980: 423).
9. It is important to note that when I grant the logical possibility of Searle's Chinese Room Scenario I am only granting that it seems logically possible that a system could be behaviourally (and functionally) indistinguishable from a mind-haver without having a mind. I am not, however, granting that a system could be physically indistinguishable from (be a duplicate of) a mind-haver without having a mind. Granting the latter gives us property dualism, while the former carries no such consequences.

I thank Marcus Arvan, José Benardete, Jonathan Bennett, John Hawthorne, Johan Küwer, Mark Scala, Robert Van Gulick and Brian Weatherson for helpful comments on various earlier drafts of this paper.

REFERENCES

Block, N. (1981). "Psychologism and Behaviorism". *The Philosophical Review*, 90, pp. 5-43.

Chalmers, D. (1996). *The Conscious Mind.* Oxford: Oxford University Press.

Dennett, D.C. (1981). "Intentional Systems". In *Brainstorms*, Cambridge, Mass: MIT Press.

Dennett, D.C. (1998). "Can Machines Think?" In *Brainchildren*, Cambridge, Mass: MIT Press,

Fodor, J.A. (1968). *Psychological Explanations*, New York: Random House.

Newell, A. and Simon, H. (1976). "Computer Science as Empirical Inquiry: Symbols and Search". *Communications of the Association for Computing Machinery*, 19, pp. 113-26.

Malcolm, N. (1954). "Wittgenstein's 'Philosophical Investigations'". *Philosophical Review*, 63, pp. 530-59.

McCartney, J. (1979). "Ascribing Mental Qualities to Machines". In M. Ringle (ed.), *Philosophical Perspectives in Artificial Intelligence*, Atlantic Highlands, N.J.: Humanities Press.

McGinn, C. (1989). "Can We Solve the Mind-Body Problem?" *Mind*, 98, pp. 349-66.

Putnam, H. (1960). "Minds and Machines". In S. Hook (ed.), *Dimensions of Mind*, New York: New York University Press.

Searle, J.R. (1980). "Mind, Brains and Programs". *The Behavioral and Brain Sciences* 3, pp. 417-457.

—— (1990). Is the Brain's Mind a Computer Program? *Scientific American*, 262, pp. 26-31.

—— (1994). *Minds, Brains and Science.* Cambridge, Mass.: Harvard University Press.

—— (1995a). *The Rediscovery of the Mind.* Cambridge, Mass.: MIT Press.

—— (1995b). "Self Portrait". In S. Guttenplan (ed.), *A Companion to the Philosophy of Mind*, Cambridge: Basil Blackwell.

Strawson, P. (1982). "Freedom and Resentment". In G. Watson, ed. *Free Will*. Oxford: Oxford University Press.
Turing, A. (1950). "Computer Machinery and Intelligence". *Mind*, 59: 433-60.
Weizenbaum, J. (1966). "ELIZA: A Computer Program for the Study of Natural Language Communication Between Man and Machine". *Communications of the Association for Computing Machinery*, 9: 36-45.

On the Rehabilitation of the Knowledge Argument

VALTTERI ARSTILA

Introduction

Arguments against one materialist theory are seldom conclusive against other materialist theories. Frank Jackson's Knowledge Argument, if it worked, would be an exception, since it directly refutes the bare minimum that Jackson himself thinks is common to all materialist theories. This bare minimum is the doctrine that "all information is physical information."

Even though one may disagree with Jackson's outcome, the intuitive force of his thought-experiment is considerable. As a result, the Knowledge Argument has been a subject of uncountable disputes. In these disputes, it has become clear that the argument is inadequate as it stands – even Jackson himself is not anymore persuaded of the argument and convinced of its outcome.[1] (Jackson 1998: 101)

In their recent article 'Mary Mary, Quite Contrary' George Graham and Terence Horgan undertake an attempt to redefine the original argument, so that its force could be better appreciated. This they try to do even they write that their "personal attractions to physical-ism are such that it is hard for [them] to accept anti-physicalism (Graham & Horgan 2000: 59)."

The purpose of this paper is to stand up for materialists and argue that the revisited Knowledge Argument, i.e. Graham's and Horgan's Mary Mary argument, does not succeed any better than the original one. This paper will begin with setting the background for the Mary Mary argument by presenting Jackson's original argument and the received view to overcome it. After that the Mary Mary argument is introduced as well as its counterargument. Finally the consequences of the presented solution will be discussed.

The Knowledge Argument

In his better-known version of the Knowledge Argument, Frank Jackson introduces a person called Mary. She is a brilliant neurophysiologist, who has lived her whole in a black-and-white room. Mary has specialised in the physiology of visual experiences, which she has studied via black-and-white books and television. She knows every significant fact there is to know about seeing the colours that objective, physical information can convey. However, despite that she knows all the physical-functional facts about color vision, she herself has never seen colors. (Jackson, 1982, p. 130)

When Mary – for some reason – leaves or is released from her room, she experiences the color-quale of the grass, sky etc. Now, the crucial question is "does she discover a new fact from her experience?" Jackson's answer is affirmative; Mary learns something about her visual experience of the world. For instance, she learns to know what it has been like for others to see red or at least what it is like for her to see red. Hence, she learns a new fact about us, and herself. (Jackson, 1982, p. 130; Jackson, 1986, p. 291)

A synopsis: the basic underlying assumption in Jackson's Knowledge Argument is that there is some knowledge about experience that can be acquired only by undergoing the relevant experience oneself. However, since Mary already knew the physical facts, these facts are not among them. Post-release Mary has learned phenomenal facts, which are something additional to purely physical information. And since this is against Jackson's definition of physicalism, physicalism is an untenable metaphysical commitment. (Jackson, 1982, p. 130)

The significance of the Knowledge Argument

The force of the Knowledge argument comes from the intuition that physicalism goes astray by misconstruing or denying the existence of the subjective and phenomenal properties of experience. Due to the generality of the argument, its consequences are significant: if the Knowledge Argument is sound, then physicalism is false. By refuting physicalism, the argument simultaneously threatens the methodological assumption that guides all scientific inquiry – not merely physics.

Physicalism also works as the basis in cognitive science and

cognitive neuroscience where scientists seem to share as their basic assumption what Jaegwon Kim has called "Alexander's dictum": to be real is to have causal powers. (Kim, 1993, p. 348) Usually scientists have regarded these causal powers to be things that have physical effects; and reciprocally, if a phenomena has no physical effects it also has no causal powers, which means that it is not real. Now the Knowledge Argument challenges this by implying that there exist real properties that are not reachable with physical descriptions, and in addition, that they have at least some causal powers, since we are able to gain new knowledge with them. Therefore, as long as the Knowledge Argument stands unassailable, the theoretical foundations for most of the empirical research remain questionable.

The types of counter-arguments

The simplest reply to the Knowledge Argument is to deny that Mary learns anything once she is released. One of the first to do this was Paul Churchland in 1985. However, most opponents of the Knowledge Argument are willing to admit that Mary learns something new. The opinions that she learns can be divided in three general categories that are in accordance with every commonly considered type of knowledge.

First, it has been said that Mary does not learn any new propositional knowledge, but know-how (skills). The first and most prominent advocates of this view are David Lewis (1990) and Laurence Nemirow (1990). They identify this know-how with the abilities to remember, imagine, and recognize experiences. This position is labeled the "Ability Hypothesis."

Second, it has also been suggested that Mary does not gain propositional knowledge or know-how, but knowledge by acquaintance. In this theory, Mary comes to know color experiences in the sense that one comes to know a city. One prominent advocate of this category is Conee (1994).

Lastly, it has been proposed that Mary learns new propositional knowledge and yet this can be explained in a way that does not refute physicalism. This position, often labelled as old-fact/new-guise theory or the mode of presentation theory, is currently the most widely held position among the critics of the Knowledge Argument. Since the subject of this paper is Graham's and Horgan's article

and they concentrate on this last view, it will also be the position discussed hereafter.

The PANIC theory

The proponents of the mode of presentation theory admit that the knowledge Mary gains when released is propositional in kind, but deny that she learns any new facts – facts that she did not know before her release. Hence these critics are willing to posit some epistemic distinction between Mary pre- and post-release, but they are certain that they have the ability to accommodate it. What they argue is that the knowledge Mary gained post-release is epistemically, but not metaphysically, progressive. As an example of a mode of presentation theory, Graham and Horgan use Michael Tye's PANIC theory, which has also independently received much attention.

Tye argues that every kind of sensory or perceptual experience has an intentional content. The intentionality of phenomenal states is secured by a covariation theory of representational content that Tye endorses. It means that the representing state causally covariates with the item represented. For example, the character of a particular color-experience is that a specific type of surface-reflectance is present. As it is the simplest explanation, Tye identifies the phenomenal character of experience with a certain sort of representational content. (Tye 1996: 93-105)

Since many or possibly all mental objects or states have intentional content (beliefs, desires), merely pointing out that every phenomenal character is a representational content does not illuminate the subject greatly. What is needed is an account of the kind of intentional contents that phenomenal characters are. This is the main aim of the PANIC theory, which Tye deploys to explain the representationalism of phenomenal character.

'PANIC' -abbreviation comes from the words 'Poised Abstract Nonconceptual Intentional Content.' This means that phenomenal contents are abstract, nonconceptual and intentional (in the sense explained earlier) contents that are appropriately poised for use by the cognitive system. (Tye 1996: 137, 143)

The requirement that contents are appropriately *poised* is to be understood as saying that phenomenal contents are in a position to have a direct impact or influence on the belief/desire system. (Tye 1996: 137-8) It is probably this part of PANIC's definition that Graham

and Horgan refer to, when they wrote that "what is psychologically significant about the PANIC properties is just the functional/representational role they play in human cognitive economy. (p. 71)"[2]

Requiring that phenomenal contents are abstract "is to be understood as demanding that no concrete objects enter into these contents" (Tye 1998: 138), i.e. our phenomenal experience can represent objects that may or may not be real. Hence, it is a question only of qualitative not quantitative identity. This feature separates phenomenal contents from (for example) belief contents, where the quantitative identity of the object of belief has an importance in determining belief content. (Tye 1996: 138)

The requirement of phenomenal content's non-conceptuality is the requirement that they can exist for their subject without the subject simultaneously possessing concepts that are a perfect match for those phenomenal contents; one can possess a phenomenal content without the corresponding concept. PANIC states do not constitute concepts merely by themselves. (Tye 1996: 138)

As Tye observes, PANIC is broadly a physical property, in the sense that there are conceptually sufficient physical conditions for its instantion. Now the position where Mary is released, is that she learns *phenomenally* what it is like to see red (red-PANIC), a state that she already knows by *physical* description. However, if she knows all the physical facts and physicalism is true, then she allegedly should know all the properties of everything. Thus, in order to meet the knowledge argument, Tye has to explain why Mary did not know the proposition "what it is like to see red" even though she did know the relevant PANICs; i.e. how Mary is able to make epistemic progress in a situation, where she knows all the facts about it.

Tye's solution to the Knowledge Argument

Tye's solution regarding Mary's epistemic progress rests on the view that we must make a distinction between fine-grained and coarse-grained facts. On a fine-grained level, the thing that differentiates two facts may be merely the mode of their presentation; they are conceptualised in different ways. For example, on a fine-grained level my belief that I drink water is different from my belief that I drink H_2O. Either of these beliefs can be had without the other. However, on a coarse-grained level these beliefs are the same, since the fact in them is fundamentally the same. So Mary is able to learn

new propositional fact on a fine-grained level, even though this fact was something she knew already on a coarse-grained level. (Tye, 1996: 172-74)

It is credible that there are more true propositions on certain facts than there are facts; the same fact can be known in various and manifold ways. Obviously in the Mary's case two important ways of knowing come via the phenomenal and physical knowledge, which however are on a coarse-grained level about the same thing.

According to Tye, in order to possess *phenomenal knowledge*, i.e. "knowing what it's like" – type of knowledge, one has to possess *phenomenal concepts*. These concepts enable one to recognize one's phenomenal properties when introspecting one's phenomenal states. Phenomenal concepts are associated with PANICs and cannot be possessed or employed without the experience that has included the PANICs related to those phenomenal concepts. Thus, one cannot "know what it's like to see" red without a foregoing experience. This also makes phenomenal concepts epistemically special and prevents them to be reduced to physical concepts. (Tye, 1996, pp. 163-64)

Tye's phenomenal concepts are - as Graham and Horgan call them – capacity-based concepts; one has to gain the capacity to use them. (Graham & Horgan, 2000, p. 66) Since physical concepts are not experience related, phenomenal concepts do not reduce to them and Mary could not possess phenomenal knowledge solely based on physical concepts, since they did not give her experiences needed to possess phenomenal concepts etc.

Now let us bring all the pieces together and briefly summarize how Tye meets the Knowledge Argument. For knowing what it is like to have an experience of *red* (phenomenal property red) it is required to possess the phenomenal concept 'Red.' Since pre-released Mary lacks the requisite concept, she does not know what it is like to see *red*. Once released, she will have experiences that enable her to gain the capacity-based phenomenal concept 'Red.' This in turn makes it possible to her to know what it is like to see *red*. Hence, Mary has gained new knowledge. However, this new propositional knowledge differs from what she earlier knew only on fine-grained level. On a coarse-grained level she has not learnt anything new. Hence, Mary's epistemic progress does not form a threat to physicalism. Graham and Horgan are willing to accept this solution (Graham & Horgan 2000: 67-68).

The Mary Mary argument

The crux of the Knowledge Argument was the possibility of phenomenal information; no physical knowledge of what goes on in the brain when one has a red experience would describe it exhaustively. In the Mary Mary argument, this emphasis has shifted to a certain degree away from phenomenal consciousness into the explanation of the psychological reactions of post-release scientist.

Graham and Horgan introduce us to Mary Mary Quite Contrary ('Mary Mary' for short), who is the daughter of Mary (Graham & Horgan 2000: 69):

> Mary Mary, like Mary, has lived in her mother's (Mary's) black and white environment and inherited her mother's physical scientific omniscience as well as monochromaticism. However unlike mother, Mary Mary has read Michael Tye's *Ten Problems of Consciousness*. As a result, she is convinced that the PANIC theory is correct. (Graham & Horgan 2000: 69)

The question of which Graham and Horgan want to make an issue of is: "When Mary Mary contemplates a post-monochromatic mental life for herself, what changes can she reasonably expect?" (Graham & Horgan 2000: 70)

First, PANIC's (phenomenal property) psychologically important feature is in the functional/representational roles it plays in human cognitive economy; these roles are, for example, the discriminatory and recognitional capacities they subserve. These Mary Mary understands already in her pre-release situation and she should therefore expect no surprises there. *Second*, since phenomenal concepts, according to PANIC theory, give us new capacities to recognise and discriminate, these new capacities are something she certainly would be anticipating and already understands thoroughly. *Third*, since phenomenal knowledge-states and equivalents are based on phenomenal concepts, she understands those states in her pre-release situation even though she cannot herself undergo them. For example, "she . . . should expect no surprises from the relevant knowledge states that constitute . . . 'knowing what it's like' to undergo color experiences." As a result, Mary Mary has good reasons to wait for learning new capacities and concepts. However, her being a rational person regarding this matter, Mary Mary has no good reason to expect anything else. (Graham & Horgan 2000: 71-72)

This, however, is not something that Graham and Horgan think. On the contrary they argue that Mary Mary will be both surprised and delighted:

> What will surprise and delight Mary Mary is (it seems to us) the unanticipated experiential basis of her concept-wielding . . . she never expected polychromatic experience to be like this. It will seem to her that this surprising, and unexpectedly delightful, aspect of her experience is both (i) distinct from the recognitional/discriminatory capacities for which it is the experiential basis, and (ii) something that has a quite distinctive character qua mental, over and above the representational/functional PANIC role that she knows about already. (Graham & Horgan 2000: 72)

However, since acceptance of Tye's PANIC theory will leave Mary Mary with no rational basis to expect surprise or anticipated delight from her post-release experience, she should conclude that Tye was wrong in claiming that possession of phenomenal concepts consists merely in certain recognitional / discriminatory capacities. (Graham & Horgan 2000: 74)

In sum, the real challenge in the Mary Mary argument is not to give an answer to the question of phenomenal information, but to give a satisfying account of why it is rationally appropriate "for her to find her post-release color experiences surprising and intrinsically delightful." (Graham & Horgan 2000: 77). And the basic problem in the PANIC theory is that it treats Mary Mary's post-release attitude as rationally inappropriate. Graham and Horgan also consider some other materialist rejoinders, which they claim are no more successful in overcoming the Mary Mary challenge than is Tye's PANIC theory. Therefore, this new version of Knowledge Argument has again brought up suspicions about physicalism.

The Mary Mary argument elaborated

The crux of Mary Mary Argument rested with the unexpected reactions after Mary Mary's release. The whole thought-experiment was based on these changes in attitude, and without them the whole argument fails. Since surprise and delight are such crucial factors here, let us elaborate more precisely on the reasons of their appearance.

As it was earlier mentioned, if post-released Mary Mary was unexpectedly surprised and delighted it might have been triggered by seeing colours. Naturally all of these properties are separate. Now should PANIC explain why seeing phenomenal property triggers delightfulness and surprisement in Mary Mary?

I claim that it is not on PANIC theory's responsibility. As Horgan and Graham mention, surprisement and delightedness are psychological states. However, at it's most simplified, PANIC theory's task is to give an account of our phenomenal characters, of our phenomenal consciousness. Of course, this also comprises giving an account of PANIC properties' functional role, which it does, for example, by explaining that PANIC properties are in a position to have a direct impact on our belief system. Nonetheless, it is not the simplified PANIC theory's task to accommodate what these impacts might be. Hence, what needs to be explained is something else than simply the psychological reactions.

Luckily, we have a candidate. Graham and Horgan argue that what seems to be the problem here is that "*real* 'knowing what it's like' is richer and more exciting than that [the capacity-based knowledge], because real phenomenal content is something over and above PANIC. (Graham & Horgan 2000: 75)" A couple of passages later they continue that phenomenal properties are "neither identical to, nor conceptually supervenient upon, properties of the kind posited by physics and by the other natural sciences. Phenomenal properties are non-physical properties with which one becomes directly acquainted in experience. (p. 75)" So to be precise, the issue in the Mary Mary argument are not the surprise and delight themselves, but that there are no rational bases for Mary Mary to be surprised and delighted by the fact that *phenomenal properties seem to be something else than PANICs.*

It is rational for Mary Mary to expect to be surprised

Even though it is not PANIC theory's responsibility to meet with the impacts of PANIC properties, its task is to explain why Mary Mary is surprised by the fact that phenomenal properties seem to be something else than PANICs. In addition, it should make it rational for her to expect to be surprised. All of this, I believe, it manages to do.

It is plausible to assume that Mary Mary, with her monochromatic vision and physical omniscientism, has been pondering what it is

like to see colours. Moreover, she probably has been doing this to her utmost; otherwise the argument would fall short.

Since Mary Mary's physiological understanding is not limited solely to colour-vision, she has, as a true PANIC-believer, probably also contemplated other phenomenal characters than the visual. For example, she has thought, via PANIC-based concepts, about her phenomenal impressions of pain and different tastes as well as their physical and functional PANIC counterparts. In these situations, she should have noticed the difference between phenomenal characters and the description of their physical realisers. After all, the former at least seem to be the simplest phenomenal properties while the latter are regarded as complex. However, according to PANIC theory, these two are the same thing. Consider now Mary Mary, as she knows that she is going to see red. Wouldn't it be possible, or even probable, that she understands that she does not yet know what it is like to see red and that this new property which she is going to see, is something she cannot yet even imagine. Is it not then rational for her in this situation to expect some kind of surprise as after all she knows that the properties she will encounter are new ones?

It is important to notice that the main argument for explaining Mary Mary's rationale for anticipation does not depend on her overall physical knowledge; she does not need to be physically omniscient in order to rationally expect a surprise. It is sufficient that she is a PANIC-believer. According to PANIC theory, phenomenal characters are second-order properties, which are realized by objective first-order physical properties. Hence, as the argument goes, second-order properties are still broadly physical properties. Now, Mary Mary should know that. In addition, she knows that she herself has never undergone states with relevant first-order properties, and therefore she has not experienced red-experiences. Once she undergoes those states, she should rationally expect to encounter new kinds of second-order properties, which will surprise her by their brand-new character. Since it is rational for her to expect surprise and the properties that surprise her are broadly physical, this presents no threat to physicalism.

This explanation of Mary Mary's rationale to expect to be surprised would not probably satisfy Graham and Horgan. What seems to be the case here, is the different kinds of view on PANICs. In the end, Graham's and Horgan's argument boils down to the

claim that PANIC-properties are something different than phenomenal character, even though this is against what Tye maintains. On their view, Mary Mary's experiential states have a functional role of the PANIC type, but "there is more to them *qua* mental, than that. There is also their *phenomenal character*, something evidently quite distinct from the various behavioral, discriminatory, recognitional and classificational capacities that jointly make up the PANIC role. (Graham & Horgan 2000: 72ff)" Consequently, as quoted above, they argue that in post-release experiences Mary Mary encounters first time phenomenal properties, which are not identical or even supervenient upon properties of the natural sciences. Phenomenal properties are wholly non-physical properties. (Graham & Horgan 2000: 75)

It is easy to agree that phenomenal character seems to be distinct from the capacities that jointly make up the PANIC role and that what is "psychologically significant about the PANIC properties is just the functional/representational role they play in human cognitive economy." However, I disagree with the authors on the claim that PANIC-properties are solely their role. On the contrary, what is *phenomenally* significant about PANIC properties is that they are phenomenal characters. PANIC properties are phenomenal characters *qua* mental, even though *qua* physicalistic framework they only seem to be some first-order properties fulfilling specific functional roles. If Graham and Horgan had not ignored the phenomenal significance of PANICs, they would have noticed that PANICs are phenomenal contents (even though their physical descriptions depict merely their functional roles). The phenomenal analysis of PANICs explains why they are second-order properties and helps one to see how they differ from first-order properties. It also makes it possible for Mary Mary to expect to be surprised particularly over the fact of how different PANICs appear to be phenomenally rather than physically – just like it was previously explained.

Something new under the sun?

Another, and independent, objection against the Mary Mary argument comes from the fact that in the elaborated problem, the situation which Mary Mary faces is already distanced too far from the original grounds. Mary Mary would face the same problems whether she has lived or not her whole life in monochromatic surroundings. Consider again post-release Mary Mary, assuming at the same time that she is, as Graham and Horgan claim, surprised

by the fact that there is a phenomenal character that does not seem to be a PANIC-property. Furthermore assume that some higher force has also told Mary Mary that physicalism and the PANIC theory of consciousness are true.

Now I claim that even though she would contemplate forever, she would not understand why PANIC properties are phenomenal properties. The reason for this is that reductionist theories of phenomenal consciousness fail to close our intuitively felt connection between phenomenal experiences and brain states – even if the theories were true. This, of course, is the gist and thrust of Joseph Levine's "explanatory gap" argument. Therefore, the purified form of the Mary Mary argument has lost the notions of phenomenal knowledge and has not brought any novel challenge to physicalistic theories by collapsing into Levine's argument.

On Tye's view also Levine's objection is based on error. Since it is outside of Graham's and Horgan's topic, I will only run through quickly how Tye has met this objection. Tye admits that there is a gap, but that this gap is caused by the different ways we conceive physical states. On the physical side the concepts we apply to physical states leave out the phenomenal feel of the state. On the other side, phenomenal concepts we apply leave out what constitutes particular experiences physically. "It is here that the gap exists, in the distance between the phenomenal and non-phenomenal modes of presentation (Tye 1996: 180)." A conceptual gap, however, does not necessarily bring along the ontological gap. Tye's commitment is therefore ontologically *minimalist* since there is only one substance. If phenomenal characters as second-order properties covary with some first-order (physical) properties, there is no need to postulate any spooky stuff in the world. Physicalism will therefore remain a viable ontological commitment.

Beyond the Knowledge Argument

Despite its enormous intuitive appeal, the Knowledge Argument loses its supporters all the time. Many of the critics have found their satisfying solution from old-fact/new guise theories, which are currently the received view. Graham and Horgan have tried to rehabilitate its importance by reconceptualizing the original argument in a way that emphasizes the psychological reactions (and their reasons) of to-be-released-scientist instead of formerly stressed phenomenal information.

The modification of the original argument, however, was not successful. First of all, the reason why Graham and Horgan had gone astray was that they did not show enough consideration for the phenomenal significance of PANICs and for the difference between first-order and second-order properties. Secondly, it turned out that at the end the intuitive force behind this new argument was not novel and one already encountered. All told, empirical researchers have no reason to feel themselves threatened: the Knowledge Argument and its newer version have not been able to disprove physicalism.

Although neither versions of the Knowledge Argument were conclusive against materialism, they however succeed in restoring some of the importance of mental phenomena. At least when the above answers are examined carefully, one notices that they bring forward a restriction in researching consciousness. This matter has often been ignored as regards the original argument, but it has to be faced with the newer version; that is, even the same conclusion can be drawn already from Jackson's argument, it is sharpened to such an extent in Graham's and Horgan's version that it is hard to bypass.

To begin with, the answer given to the original knowledge argument leans on the idea that there are concepts that cannot be had without experiences containing relevant PANICs. Hence, our phenomenal knowledge is not dependent solely on concepts, but ultimately it depends on the experiences that make it possible for us to have and form phenomenal concepts. According to the theory defended in this paper, phenomenal properties (PANIC properties) are second-order properties, which covariate reliably with some first-order properties. However, as the answer to the Mary Mary argument illustrates, even though the second-order properties covary with first-order properties, PANICs cannot be understood even as properties beforehand from the first-order property description.

The important consequence of this is that one cannot examine phenomenal consciousness only with physical terms. No matter how strongly and clearly phenomenal properties covariate with first-order properties, we cannot know what kind of second-order property the first-order properties have instantiated without at least the first experience of phenomenal properties! From an empirist's point of view this means that decent research on phenomenal consciousness

cannot be based solely on physical properties. Researchers need to take phenomenal consciousness seriously even though the properties within phenomenal consciousness are still broadly physical properties.

Acknowledgement

The author is supported by the Finnish Cultural Foundation.

REFERENCES

Churchland, P. M. (1985). Reduction, qualia, and the Direct Introspection of Brain States. *The Journal of Philosophy, LXXXII*(1), 8-28.

Conee, E. (1994). Phenomenal Knowledge. *Australasian Journal of Philosophy, 72*(2), 136-150.

Graham, G. & Horgan, T. (2000). Mary Mary, Quite Contrary. *Philosophical Studies, 99*, 59-87.

Jackson, F. (1982). Epiphenomenal Qualia. *Philosophical Quarterly, 32*, 127-36.

Jackson, F. (1986). What Mary Didn't Know. *The Journal of Philosophy, 183*, 291-295.

Jackson, F. (1998). *From Metaphysics to Ethics, a Defence of Conceptual Analysis*: Oxford: Clarendon Press

Kim, J. (1993). *Supervenience and Mind*. Cambridge, Mass.: Cambridge University Press.

Lewis, D. (1990). *What Experience Teaches*. In W.G. Lycan (Ed.), *Mind and Cognition* (pp. 499-518). Cambridge, Mass.: Basil Blackwell.

Lloyd, D. (1997). Consciousness and its Discontents. *Communication & Cognition, 30*(3/4), 273-284.

Metzinger, T. (1995). Faster Than Thought. In T. Metzinger (Ed.), *Conscious Experience* (pp. 564). Paderborn: Ferdinand SchÏningh.

Nemirow, L. (1990). Physicalism and the Cognitive Role of Acquaintance. In W. G. Lycan (Ed.), *Mind and Cognition* (pp. 490-8). Cambridge, Mass.: Basil Blackwell.

Tye, M. (1996). *Ten Problems of Consciousness, A Representational Theory of the Phenomenal Mind*. Cambridge, Massachusetts: MIT Press.

NOTES

1. Two quantitative different objects that are exactly alike regarding their perceivable properties can be substituted for each other without altering the phenomenal character they give rise to.
2. The requirement of the rational basis in important here, since in the latter part of their article, Graham and Horgan considered the materialist rejoinder that claims Mary Mary to be in a position to predict that she will be surprised. Momentary they are willing to agree with this. Furthermore, they are willing to

grant that Mary Mary's scientific omnipotence enables her to know the neurophysiological basis of surprisem and delight, as well as to know "in advance any post-release changes in her own neurophysical processes that would be relevant to the realization of psychological states like surprise and delight (Graham & Horgan, 2000, p. 76)." However even this does not help physicalists and the problem remains since the real challenge is to give an account of why it is *rational* for Mary Mary to anticipate surprise and delight.
3. Of course it would not do to argue that PANICs have already made it possible to Mary Mary to surpass apparent differences in phenomenal pain-experiences and their physical descriptions, since then one would have to explain why she could assume otherwise regarding visual properties? Without plausible reasons for this unequal treatment of different PANIC-properties, this objection fails.

Psychological Reduction, the Autonomy of Psychology, and Disjunctive Causal Properties

SVEN WALTER

In this paper I discuss a dilemma for non-reductive physicalism, currently the received view in the philosophy of psychology, that was formulated by Jaegwon Kim in 1992. Accoring to Kim, non-reductive physicalism is a myth: *Either* one must advocate the wholesale reduction of psychology into physical science reducing psychological properties to *disjunctions* of their possible physical realizers, *or* one must advocate the sundering of psychology in distinct fields (as many as there are physical types to which the various physical realizers of a psychological property belong), each one of which is directly reducible to physical science (Kim's label for this is *local reductionism*).

I argue that when faced with this dilemma, we should opt for the first horn (contrary to Kim who takes the second route and defends local reductionism). This, however, requires an account of what *disjunctive properties* are and how they can be legitimate causal properties. I will try to give an account of disjunctive causal properties that turns on a distinction between two kinds of multiply realizable higher-order properties like psychological properties.

Finally, I will consider the question where we stand with respect to the autonomy of psychology if we really advocate the reduction of psychological properties to disjunctive physical properties. I will urge that even though psychology is *in-principle* reducible, it still is *in-fact-autonomous*, and that is all autonomy we need. Psychophysical Reduction, the Autonomy of Psychology, and Disjunctive Causal Properties*

To think that one can be a serious physicalist and at the same

*I apologize for the assaults on the English language. My thanks to Lousie Antony and Carl Gillett for various helpful discussions.

time enjoy the company of things and phenomena that are nonphysical, I believe, is an idle dream. Reductive physicalism saves the mental, but only as a part of the physical (Jaegwon Kim 1998)

SETTING THE SCENE

In a number of papers during the last fifteen years, Jaegwon Kim has vehemently attacked *non-reductive physicalism* as a possible account of the relationship between psychological properties (intentional as well as phenomenal) and physical properties, particularly neurophysiological ones. In spite of his criticism, non-reductive physicalism still is something like the 'received view' in current philosophy of mind and philosophy of psychology. One of the main reasons for the predominance of non-reductive physicalism seems to be the fact that it promises to allow for the *autonomy* of psychology and other special sciences (by assuming psychological, aerodynamical, genetical etc. properties as the *irreducible* and hence *distinct* subject matter of psychology, aerodynamics, genetics etc.—this is the *non-reductive* part of non-reductive physicalism) while at the same time enjoying the advantages of full-fledged *physicalism* (by arguing that psychological, aerodynamical, genetical etc. properties are *realized by* or *supervenient on* or *dependent on* basic physical properties and hence causal properties in good standing—this is the *physicalistic* part of non-reductive physicalism). To put it in a slightly different way: According to non-reductive physicalism the autonomy of psychology and the other special sciences stands and falls with the irreducibility of the properties that allegedly form their distinct subject matter.

The irreducibility of psychological properties in turn seems to be assured mainly by two arguments: First by Donald Davidson's argument for the *anomaly of the psychological* to the extent that psychological properties[1] are conforming to some rules of *rationality* and principles of *normativity* that have 'no echo in physical theory'—a fact that seems to rule out any reductive relationship between psychological and physical properties. Second, and more important, the irreducibility of psychological properties (and with it the autonomy of psychology) seems to be guaranteed by the fact that psychological properties are *multiply realizable* in a wide variety of physical properties. The *multiple realization argument (MRA)* was first brought to our attention by Putnam in (Putnam 1967) and has later

been adopted by, among others, Jerry Fodor. The thesis of the multiple realizability of psychological properties says that creatures of different species than humans (Martians, dogs etc.) may instantiate our psychological properties even though in them those properties are differently physically realized than in us human beings. If this is correct, psychological properties can't be *identical* with any of the physical properties by which they are realized. Non-reductive physicalism has an easy way to accommodate the multiple realizability of psychological properties, since its central claim concerning the relationship between psychological and physical properties is just that of *supervenience* (or *realization*) and not that of *identity*. One of its his major rivals, *reductionism*, however holds that psychological properties are *type-identical* with physical properties (see Smart 1959 for the classic exposition). Therefore, if psychological properties are indeed multiply realizable, reductionism is excluded.

For a long time it seemed as if non-reductive physicalism, supported by *MRA*, carried the day compared to reductionism. Consequently, reductionism has had a hard time.[2] Nowadays "consensus holds that reductionism is dead" (Bickle 1998: 1), which is a rather generally acknowledged attitude that being a reductionist "is a bit like being a logical positivist or a member of the Old Left: an aura of naivetÈ hangs over such a person" (Kim 1989: 32). Recently, however, Jaegwon Kim turned the tables by formulating a dilemma for non-reductive physicalism that unveiled a rather surprising insight (Kim 1992): The multiple realizability of psychological properties does not *ensure* the autonomy of psychology but on the contrary seems to *deprive* psychology of its autonomy!

According to Kim, anyone who advocates non-reductive physicalism on the basis of the multiple realizability of psychological properties faces the following predicament: One must *either* advocate the wholesale reduction of psychology into physical science since multiply realizable psychological properties can be reduced to physical properties by identifying them with *disjunctions* of their possible physical realizers, thereby allowing for psychophysical laws expressing the *nomic coextensivity* of psychological and physical properties. *Or* one must advocate the sundering of psychology in distinct fields (as many as there are physical types to which the various physical realizers of a psychological property belong), each one of which is directly reducible to physical science (Kim's label for this is *local reductionism*).

The plan for the remainder of this paper is as follows: In section 2 I will consider the *MRA* in detail and show how and why Kim's dilemma for non-reductive physicalism arises. Furthermore, I will briefly discuss both horns of the Kim-dilemma. While Kim himself has chosen the second horn, thereby embracing local reductionism, I will urge that in the mind-body case the first horn, the *disjunction-strategy*, is more appropriate. In section 3 I will discuss some of the objections to the legitimacy of *disjunctive properties* and I will argue that some disjunctive properties indeed are legitimate causal properties 'in good standing'. I will argue that there are some multiply realizable properties that are such that an identification with disjunctive properties is indeed excluded but that there are some cases in which such an identification is legitimate. The central claim which will be supported by empirical data is that the mind-body case belongs to the latter category. Finally, in section 4, I will discuss the question where we stand on the present account—the reduction of psychological properties to disjunctions of their physical realizers—with respect to the autonomy of psychology, and I will argue that we still have enough autonomy.

The Multiple Realization Argument

According to *MRA*, psychological properties cannot be type-identified with physical properties since psychological properties are realizable in—as it is often put—*a wide variety* of basic physical properties (later on we will see that the 'wide variety'-part is largely a fairy tale). Here is *MRA* in a more formal manner:

(1) There is at least one psychological property *M* for which there is a (possibly infinite) set of sufficiently distinct physical properties $\{P_1, ..., P_n\}$ (n > 1) such that each P_i d$\{P_1,..., P_n\}$ is *sufficient* for *M*,[3] no single P_i is necessary for *M* and $\{P_1, ..., P_n\}$ *exhaust* the set of *M*'s possible realizers.

(2) The only possible candidates for an *identification* with *M* are P_1 through P_n, i.e. if *M* can be reduced at all it can be reduced to a P_i.

(3) *M* cannot be identified with one single $\underline{P_i}$ since it is not even *nomically coextensive* with one single $\underline{P_i}$ (since *ex hypothesis* $P_1, ..., P_n$ are *mutually distinct*).

(c) There is no physical property with which *M* is nomically coextensive and therefore none with which *M* is identical. Hence there is no reduction of M̲ to physical properties.

Non-reductive physicalism typically assumes that psychological

properties *supervene* on physical properties, where *supervenience* is construed as a relation between two families of properties, both of which closed under the Boolean operations of complementation, disjunction and conjunction. There are various non-equivalent versions of supervenience, but for our concerns the following rough idea of *strong supervenience* suffices: A family of (e.g. psychological) properties Y supervenes on a family of (e.g. physical) base-properties F just in case the fact that something has a particular F-property *P* is sufficient for it having the particular Y-property *M*, i.e. just in case anything with that Y-property *M* has a F-property P_i such that every object with that F-property P_i must have that Y-property *M*:

(*STRSUP*) Necessarily, for every event x and every psychological property *M* of x, x has some physical property P such that necessarily whatever has *P* has *M*.

Formally: o "$M \in$ Y "x (Mx $\supset \exists P \in$ F ($\underline{P}$x $\wedge$ o "x ($\underline{P}$x $\supset \underline{M}$x)))

Interestingly, one can show that *STRSUP* entails that for any psychological predicate M there is a physical predicate P^* such that P^* and *M* are *necessarily coextensive* (Kim 1978, 1984):

(*NC*) o "x (Mx $\equiv P^*$x)

If there is a legitimate property denoted by P^* it can't be one of the P_i for the reasons given above. Notice moreover that the second premise of *MRA* only holds that M cannot be identified with one of its physical realizers P_i, not that it cannot be identified with any physical property *at all*. This means that in order to deny premise (2) and save reductionism from *MRA*, we would have to find a physical property different from P_1, ..., P_n with which *M* can be identified. So why not take P^*—whose existence is guaranteed by *NC*, which in turn follows from *STRSUP*? The problem clearly is the nature of the predicate P^*. P^* is a disjunctive predicate that has the form $P_1 \vee ... \vee P_n$. Forming disjunctive predicates is not so much of a problem, but the important question is whether there is a *legitimate* property denoted by disjunctive predicates: Are there *disjunctive properties*? Can we take P^* to be the property defined by *disjoining* the *properties* P_1, ..., P_n? If so, this disjunction would be nomically coextensive with our psychological property *M* and premise (2) of *MRA* would be false.

If disjunctive properties are legitimate properties and the disjunction P_1, ..., P_n is *exhaustive*, P^* will be necessary and sufficient

for *M*. This is the way we could archive the *bridge-laws*—statements expressing relations of nomic coextensivity between the kind predicates of psychology and the kind predicates of physical theory—required for the classic Nagelian reduction of psychology to physical theory:

(*BL*) o "x (Mx $\equiv$ ($\underline{P}_1$x $\vee$ ... $\vee$ P_nx))

However, whether there really are such 'metaphysical hulks' as *disjunctive properties* is a tendentious matter and many will vehemently argue that there aren't. I will come back to that issue in the later parts of this paper. Let us now turn to the second horn of the dilemma, the one that Kim himself embraces: *local reductionism*.

The first premise of *MRA* contains an existential claim. However, this claim can be denied, at least according to some authors (Lewis 1966, 1972; Armstrong 1968; Kim 1998). Premise (1) states that there is *one* psychological property *M*, e.g. pain, instantiated in creatures of various species which has (possibly infinitely many) sufficiently distinct physical realizers, depending on the kind of species. One might however argue that this is an illusion: There is not one single psychological property that is multiply realized by different physical properties but there are *different psychological properties* in humans, Martians, dogs etc., each of which is identical to a physical property, even though all are picked out by the same concept, viz. 'pain'. Consequently, the phenomenon referred to by non-reductive physicalism would be not so much multiple *realization* of psychological *properties*, but would be properly described as the multiple *reference* of psychological *terms*. According to this position (originally independently developed by Armstrong and Lewis), psychological concepts and psychological terms are *non-rigid designators*.

If this is true, the (erroneously) so-called multiple realization of psychological properties in different kinds of creatures can be accommodated easily and naturally by psychophysical reductionism: A psychological concept and the term expressing it simply non-rigidly attribute *different properties* to creatures of different species: it non-rigidly attributes a certain neurophysiological property to Earthlings and a different physical property to Martians etc., so that we end up with many different kinds of pain-properties (*human pain, Martian pain, dog pain* etc.) all of which are *locally reducible* to basic physical properties. What we don't have, however, is any such single property like *pain*. We have to give up *M* as a *genuine higher-order psychological*

property. We ultimately only retain the expression '*M*' or the concept *M* by which we, as it turns out, pick out various distinct properties in the base domain depending on the kind of system, species or structure in question, thereby allowing for so-called species- or system-specific reductive bridge-laws:

(SSBL) "x ($\underline{S}$x ⊃ (*M*x ≡ *P*x)), where *S* is a species

According to this kind of *local reductionism*, the position Kim currently defends himself, the existential claim in premise (1) is false and this again allows us to reject the conclusion of *MRA* and save reductionism.

The Kim-dilemma for non-reductive physicalism can thus be summarized as follows: *Either* psychological properties can be reduced to physical properties *via* the disjunction-strategy, *or* psychology falls apart in various sub-disciplines (human psychology, Martian psychology, canine psychology etc.), each of which is directly (locally) reducible to physical science. Part of what I want to do in the remainder of this paper is to consider the question whether Kim is well-advised in dismissing the first horn of the dilemma and instead opting for *local reductionism*. Even though local reductionism admittedly is a kind of *reductionism*, it still is a version of *eliminative reduction* in which the reduced entity—the psychological property *M* in our case—does not survive the process of reduction; eliminative reduction is a kind of reduction that rids our ontology of the reduced entities. Personally, I do not want to give up my vast conviction that I, my girlfriend, my dog and my girlfriend's cat do have something in common—that we literally do have a *property* in common—when we experience a pain in our leg. According to local reductionism, however, this is an illusion. Therefore I think it might be worth considering whether the disjunction-strategy can be made to work and this comprises the question whether there are such things as *disjunctive properties*.

Disjunctive Causal Properties

In one of his early discussions of *MRA*, Hilary Putnam addressed the possibility of identifying a higher-level (psychological) property with the disjunction of its lower-level (physical) realizer properties, but readily dismissed it as the last act of desperation of reductionists and concluded that "this does not have to be taken seriously"

(Putnam 1967). I don't think we should dismiss disjunctive properties that willingly, but admittedly the disjunction-strategy *prima facie* is at least questionable. In what follows, I will concentrate on the three most pressing problems with disjunctive properties.

The first difficulty is formulated paradigmatically by David Armstrong in the following passage:

> Suppose that a has *P* but lacks $\underline{Q}$. The predicate '*P* ∨ $\underline{Q}$' applies to a. Nevertheless, when a acts, it will surely act only in virtue of its being $\underline{P}$. Its being $\underline{P}$ or $\underline{Q}$ will add no powers to its arm. (Armstrong 1978, 20)

Similarly, Terry Horgan wrote:

> The real problem with disjunctive or infinitely disjunctive properties . . . is that they are not causal properties: on any occasion when such a property is instantiated, the 'causal work' will be done not by the whole property, but only by whatever specific *disjunct* is instantiated on the given occasion. (Horgan 1996, 595)

I have to admit that it is not quite clear to me what that objection amounts to. On the most widely held construal of causation, causation is a relationship between *events*[4] and one famous conception of events is Kim's conception of events as *property exemplifications*. If we accept Kim's position that events are property exemplifications (instead of adopting, say, a Davidsonian theory of events as structured particulars), 'whole properties' never do any 'causal work': What is causally efficacious on each occasion of a causal transaction are property *exemplifications*, i.e. instantiations. I suspect that what worried Terry and David Armstrong is the fact that those causally efficacious instantiations 'normally' belong to the same physical type, whereas in the case of disjunctive properties they belong to different types—at least from point of view of basic physics. But if we regard disjunctive properties (composed out of basic physical properties as disjuncts) as legitimate physical types, this problem should vanish. Whether, however, we can regard disjunctive types as legitimate causal/natural/physical types is a question to which we turn now.

The second difficulty with respect to disjunctive properties is already anticipated in some discussions of functionalism (e.g. Block 1980) when it is asked what all instances of a certain functional

property have—given the differences in their physical make-up—in common in virtue of which they all count as instances of this property. Armstrong has put it as follows:

> Disjunctive properties offend against the principle that a genuine property is identical in its different particulars. Suppose a has property *P*, but lacks *Q* while b has *Q* but lacks *P*. It seems laughable to conclude from these premises that a and b are identical in some respect. Yet both have the 'property' *P* or *Q*" (Armstrong 1978, 20).

This objection is related to the third and probably most pressing problem for disjunctive properties in way that will become clear when we proceed. This third problem is that it is not clear whether generalizations that refer to disjunctive properties express genuine *laws* since it is not clear whether they fulfill the necessary condition of being *confirmed* by their instances. This in turn has to do with the fact that disjunctive properties seem not to be *projectible* and, therefore, not to designate genuine *causal/scientific kinds* or *types* (see Fodor 1974). I assume a familiarity with the relevant arguments, so I can be very brief. Suppose we have a generalization of the following form

(4)"x (Mx $\supset$ Rx),

and we observe a large number of *M*s that are *R*. Suppose that *M* is the psychological property that we want to identify with the disjunctive physical property $P_1 \vee P_2$. Our observations clearly count as *positive instances* of the following regularity:

(5)"x ((P_1x $\vee$ P_2x) $\supset$ Rx),

but they do not *confirm* it since it might turn out that in all the instances of *M* we observed, *M* was realized by P_1 which does not rule out the possibility of P_2s that are not *R*. In general, evidence for an *A* being a *C* is not evidence for a *B* being a *C*. Moreover, if disjunctive properties are not projectible, we do not just have reason to believe that they cannot be referred to in laws or lawlike generalizations, we then also have good reason, as has been pointed out forcefully by Fodor (1974), to assume that they do not designate *causal/scientific kinds*.

How can we respond to that objection on behalf of disjunctive properties as legitimate causal properties/kinds? If we consider the

second objection—that the instances of disjunctive properties are not 'identical in some respect', we see that this might well be true for Armstrong's favorite disjunctive property *is a raven or a writing desk* but that this is by no means true of *all* disjunctive properties. Granted, the individual realizers of Armstrong's property are *causally* so much *heterogeneous* that they have nothing *relevant* in common.[5] But that is not sufficient to score a point against disjunctive properties in general since Armstrong would have to show that *no* disjunctive property *can be* such that its instances are identical in some respect.

The same argument applies to the third objection. What is it for a property to be projectible? The most plausible answer is that in order for a property to be projectible, it must guarantee some degree of *similarity* among the individuals possessing that property. And this similarity, it can be plausibly argued, is best viewed as *similarity in causal powers*, i.e. *causal homogeneity*. Once again, *is a raven or a writing desk* turns out not to be not projectible since there is no causal homogeneity among the individuals falling under the predicate 'is a raven or a writing desk'. But again Armstrong has *not* shown that there *cannot be* disjunctive properties that are such that they guarantee some (sufficient) degree of causal similarity among their instances.

To fully make my point, however, I should at least sketch an example of what I have in mind when I talk about disjunctive properties that 'are identical in some respect' or 'guarantee some degree of causal homogeneity'. To begin with, there is something about functionalism that strikes me as quite surprising since I've heard of functionalism for the first time in my very first undergraduate seminar on the philosophy of mind: The assumption—that most functionalists seem to find *totally innocent*—that the *functional role* something plays is totally independent of its *physical make-up*:

> Functionalism . . . often pretends to invoke a perfectly clear distinction when it talks of function and its multiple realizations. I have always found this surprising. In most versions of functionalism 'function' means 'causal role'. But causal roles derive from a material's causal properties. (Bieri 1995: 53)

It seems to me that there is a much more intimate connection between what basic physical properties a thing has and what functional properties it is capable of having and it is by no means so

that any functional property can be multiply realized in a *wide variety* of physically *absolutely diverse* realizers. Rather it seems that there are radical differences among multiply realizable properties concerning the degree to which 'multiply' really means 'physically diverse'. Some multiply realizable higher-order properties—like *paper weight* or *I can throw it after my cat if she nerves me*—do not impose many *causal constraints* on the first-order physical properties that can possibly realize them. Consequently physical properties can qualify relatively easily as an occupier of the causal role definitive for predicates like 'paper weight' or 'I can throw it after my cat if she nerves me'. That's why quite a lot of otherwise (i.e. from a physical point of view) *causally* very *heterogeneous* properties form the set of realizer properties. If we were to identify such a multiply realizable property with the disjunction of its realizers, we would run into precisely the difficulties pointed out above.

On the other hand, however, there are multiply realizable higher-order properties—like e.g. *combustible, reflecting, cuts glass*—that impose much more specific causal conditions on the first-order physical properties that potentially realize them. Consequently, much less physical properties fulfil those constraints, thereby qualifying as realizers of the multiply realizable property in question. Moreover, since they have to fulfil *causal* constraints, those that *do* qualify are *much more causally homogeneous* then the realizers of, say, *paper weight*. In *those* cases we might well identify the multiply realizable higher-order property with the disjunction of its physical realizers and archive a legitimate, *causal disjunctive property*.[6]

To put it a bit more generally: I think it is a very weird idea that any functional or causal role can be played by any or even by a wide variety of physical properties: In order to count as something that has the property *cuts glass*, a given thing has to have a much more specific set of basic physical properties than in order to have the property *I can throw it after my cat if she nerves me*. Perhaps there might even be extreme cases in which the causal conditions imposed on the possible realizer properties are so highly specific that only a very few (perhaps even only *one*) physical properties satisfy it.[7]

We are now in a position to see that there *are* cases in which Fodor's worries about the *projectibility* of disjunctive properties and about the causal/scientific status of disjunctive kinds are unwarranted. Those are precisely the cases in which the multiply realizable higher-order property is such that it imposes *strong causal*

conditions on the first-order physical properties that possibly count as its realizers. Consequently, the higher-order property in question isn't literally 'multiply' realizable in a *wide variety of physically* (and hence *causally*) *heterogeneous* properties but just in a small number of properties that are similar in causal powers. This insight also takes the edge off of Armstrong's objection that the instances of disjunctive properties are not 'identical in some respect': Some of them indeed *are*—not those of Armstrong's *is a raven or a writing desk*, but, e.g., those of the disjunctive property that is archived by disjoining the first-order physical realizer properties of *combustible, reflective* or *cuts glass*. Contrary to what Armstrong says, it is not at all 'laughable to conclude' that the first-order physical properties that realize those three higher-order properties are 'identical in some respect'. They really are!

If the account sketched here is correct, there *are* multiply realizable special science properties that can be identified with disjunctive properties (composed out of their physical realizer properties as disjuncts) that are perfectly well-behaved, genuine, projectible, causal properties since there is a sufficient degree of causal homogeneity among their realizer properties. Clearly however, there are also those special science properties for which this disjunction-strategy does *not* work, viz. the ones that do not pose specific causal constraints upon their physical realizer properties. In those latter cases, I am afraid, Kim's *local reductionism* might be the appropriate move: We simply eliminate the higher-order property in favor of various first-order properties which, in *those* cases, after all isn't that bad since causally so many heterogeneous properties like the ones we are talking about now do not constitute causal, scientific or natural kinds *anyway*, so what good would it be to retain them?[8]

Unfortunately I have to admit that I do not have any clear-cut-criteria available that allow us to distinguish the two cases in general. But I do not just see no reason to suppose that the mind-body case belongs to the latter sort (at least there seems to be no reason offered in the relevant literature), there seems to be even empirical evidence that the idea of psychological properties that are *multiply realizable* in a wide variety of physically diverse first-order properties is a *philosopher's fiction* (even though I can't go into details here):

> The claim that psychological states are . . . multiply realized is usually argued on fairly intuitive grounds. It is widely assumed that different individuals . . . can be in the same psychological

state. . . . Yet, it is well known that even within a species brains differ. Even within an individual over time there are differences. . . . This line of reasoning appears plausible, but a proper examination of neurobiological and cognitive neuroscience practice will show that the claim that psychological states are in fact multiply realized is unjustified, and that what is usually taken to be evidence for it, is not. (Bechtel & Mundale 1999: 176-7)

Furthermore it seems that there is recent evidence in the neurosciences that once one directs one's attention to the sub-neuronal level, the phenomenon of multiple realizability is not warranted:[9] There seems to be evidence that what goes on on a sub-neuronal level in humans, dogs etc. when they, say, feel pain is relatively the same. Further empirical research in this area might well substantiate the theoretical considerations concerning multiple realization and disjunctive causal properties set forth so far.

Let us briefly turn to an objection. I urged that some multiply realizable properties are such that they guarantee a sufficient degree of causal similarity among their physical realizer properties while others don't. Clearly, the phrase of a 'sufficient degree of causal similarity' is crucial to my purposes. The problem with that phrase is that—to modify Hume's remark—everything is similar to everything. So one might argue that even properties like *paper weight* or *I can throw it after my cat if she nerves me* guarantee some degree of *causal similarity* among their first-order realizer properties since all the things that have those properties are similar in some respect: they all are extended objects, have a mass greater than that of air etc. First of all, what really matters for causal homogeneity among different things seems to be a question of their *intrinsic* properties, whereas the standard properties alluded to in trivial similarity cases (everything is similar to everything) are *extrinsic* properties as far as I can see.

Moreover, let us consider an illustrative example from the history of philosophy. In early analytical philosophy, one famous attempt to define *similarity classes* or *kinds*, adapted from Carnap, was the following: A set of things is a kind if all the members of the set are more similar to one another then they all are to any one thing outside the set. In other words: Every non-member differs more from some member than that member differs from any member. Alas, this definition succumbed to what Goodman has called the *difficulty of imperfect community* in his criticism of Carnap (Goodman 1966: 163f):

Consider the set of all red round things, all red wooden things and all round wooden things; this set fulfils Carnap's definition of a class, but it is not what we mean by a kind, since "it admits yellow croquet balls and red rubber balls while excluding yellow rubber balls" (Quine 1969, 121).

It is very illustrative for our purposes that no one argued against Goodman that Carnap's definition was *right* since all the things mentioned above *are* similar in some respect: They have an extension, they have a mass etc. We have a fairly good intuitive sense of which things or properties are causally similar (i.e. which sets of things exhibit a 'sufficient degree of causal similarity') and which are not. According to this intuitive judgement, I claim, the realizers of properties like *paper weight* or *I can throw it after my cat if she nerves me* count as less causally similar then those of *cuts glass, reflecting* or *I can light my barbecue with it.* Maybe, at some time in the future when much more progress in neuroscience has been made, they will tell us that the physical realizer properties of psychological properties (or at least some of them) are causally sufficiently similar so that nothing prevents us from identifying them with the disjunctions of their physical realizer properties. But so far, I think, we are entitled to conclude "that the standard arguments for multiple realizability of mental properties offer nothing against [the] reduction of higher-order properties to lower-order disjunctive properties" and that therefore "the multiple realizability arguments do not establish the metaphysical point that mental properties are distinct from any physical properties" (Antony 1999, 9).[10]

Admittedly, this is outright, radical *neuronal chauvinism*, and I still do not feel very comfortable with the idea of 'metaphysical hulks' such as disjunctive properties. But if the only alternative is *local reductionism*, which requires giving up my vast conviction that my dog, my girlfriend, my girlfriend's cat and I have literally the same *property* in common when we *feel pain*, I stick with neuronal chauvinism and disjunctive properties.

The autonomy of psychology

If the account sketched in this paper is correct and we can identify (at least some) higher-order special science properties with

disjunctive causal properties, viz. with the disjunctions of their first-order physical realizer properties, would that mean that we have to give up on the autonomy of psychology since from the viewpoint of the special sciences their autonomy is threatened (and ultimately vanishes) if their 'distinct' subject matter can be reduced to the ordinary subject matter of physical sciences? In the remainder of this paper I will argue that even though the disjunction-strategy (in some cases) provides us with an *in-principle* possibility of *ontologically reducing* the subject matter properties of psychology and other special sciences, there remains some *in-fact-irreducibility* for *epistemological* reasons. This *in-fact-irreducibility*, I urge, accounts for some kind of *epistemological autonomy* that should—must—be sufficient for our concerns.

To begin with, as briefly noted above, what matters for the reduction to disjunctive properties is not the *number* of the realizer properties that fulfil the causal constraints imposed by the multiply realizable higher-order property but the strictness or specificity of those causal constraints. So it might very well be that in a concrete case we are still dealing with an *extremely large number* of (causally relatively homogeneous) physical properties—so the disjunction might be far longer than any of us could ever fathom in a lifetime, if not even infinite. Furthermore, the disjunction as a whole might—despite its causal homogeneity—appear to be *unprincipled* for the *untutored* since the individual properties will not reveal any significant commonalties other than on the level of *advanced neuroscience*. Those two factors lead to the fact that in concrete cases—even though we might be able to *verify* the disjunctive property very quickly, i.e. to determine that a given object *does* have the disjunctive property in question— we probably never will be able to *falsify* it for certain, i.e. to *definitely rule any object out*. This will increase the practical difficulty of determining whether a particular given object has the disjunctive physical property in question or not. Therefore it will be much more easy—in fact it will be the only *practicable* way—to retreat to the generalizations and predicates (or properties) of the *special sciences* again. This is the reason we ultimately end up with an *in-principle-reducibility* of psychological properties but with a practical and *in-fact-autonomy* of psychological generalizations, predictions and explanations.

In-principle reducibility clearly seems to be enough reducibility to satisfy the demands of classical reductionism. But, and this is the

much more important question, is *in-fact-autonomy* enough autonomy to serve the demands of the special sciences?[11] I don't have much to say about that issue here. To determine whether *epistemological* or *in-fact-autonomy* is enough is the task of the special scientists. Suffice it here to say that I do not see how we can give them more of autonomy.

NOTES

1. Davidson explicitly restricts his argument to *intentional* psychological properties.

2. Note that nothing so far essentially hinged on the assumption that we are dealing with *psychological* properties. The same considerations apparently apply to all other 'higher-level' *special science* properties in relation to underlying 'basic' or 'lower-level' *physical* properties.

3. Actually, one would have to talk of *instantiations* of P_i that are sufficient for the *instantiation* of *M* but I will avoid the talk of instantiations for brevity's sake throughout this paper.

4. Or, if you wish, between events *under a description* or events *qua* belonging to a certain type, it doesn't matter for the present purposes.

5. Even though they certainly *do* have something in common (e.g. they have a mass etc.).

6. Note that this provides us with a solution to the long discussed question how to handle 'dispositional properties'!

7. The important point is not the concrete *number* of realizer properties that satisfy the causal conditions imposed but that those conditions certify that there is some (*more* or *less*) degree of *causal homogeneity* among them; the more specific the conditions are, the greater is this amount of *causal homogeneity* and if it is very high, there still might be an enormous *number* of properties that constitute the relevant set of—causally homogeneous—properties.

8. Here I am uncritically assuming that Kim was right in his discussion with Fodor (see Kim 1992, Fodor 1974) that if the set of physical realizer properties of a given multiply realizable higher-order property does not constitute a causal/scientific kind, than neither does the higher-order property.

9. I talked about this issue to John Bickle (University of Cincinnati) privately, but I wasn't able to find any published work on it yet.

10. As the quoted passage continues, however, Louise adds that they do establish an epistemological point about what modes of description are and are not practicable within human epistemic enterprises and subsequently argues that this epistemological point has metaphysical implications which causes her, if I understand her correctly, to embrace a version of non-reductive physicalism. Obviously, I don't agree in this last point.

11. At some point in the past the opponents of reductionism, the advocates of non-reductive physicalism, claimed to have found a relationship between psychological and physical properties which promised to be strong enough to give them *dependence*, while at the same time weak enough to avoid outright *reduction*—*supervenience*. Ultimately, however, it turned out that every notion of supervenience

that was strong enough to give them dependence (viz. *strong* supervenience) was also strong enough to yield psychophysical reductions, and that every notion of supervenience that avoided those reductions (viz. *global* or *weak* supervenience) was too weak to yield the dependence of the psychological on the physical. We are somehow having the opposite problem with the disjunction-strategy: Psychophysical reduction to disjunctive physical properties must be a relationship strong enough to count as reduction, but weak enough to preserve autonomy.

REFERENCES

Antony, L. (1999): Multiple realizability, projectibility, and the reality of mental properties. *Philosophical Topics*.

Armstrong, D.M. (1968). *A Materialist Theory of the Mind*. London: Routledge.

Armstrong, D.M. (1978). *A Theory of Universals*. Cambridge: Cambridge University Press.

Bechtel, W. & Mundale, J. (1999). Multiple realizability revisited: Linking cognitive and neural states. *Philosophy of Science* 66: 175-207.

Bickle, J. (1989). *Psychoneural Reduction: The New Wave*. Cambridge, Mass.: MIT Press.

Bieri, P. (1995). Why is consciousness puzzling? In T. Metzinger (ed.), *Conscious Experience* (pp. 45-60). Paderborn: Ferdinand Schoningh/mentis, 45-60.

Block, N. (1980). What is Functionalism? In N. Block (ed.), *Readings in the Philosophy of Psychology* (pp. 171-84). Cambridge, Mass.: Harvard University Press.

Fodor, J.A. (1974). Special sciences. *Synthese*, 28, 97-115.

Goodman, N. (1966). *The Structure of Appearance* (2nd ed). New York: Bobbs and Merill.

Horgan, T. (1996). Kim on the mind-body problem. *British Journal for the Philosophy of Science*, 47, 579-607.

Kim, J. (1978). Supervenience and nomological incommensurables. *American Philosophical Quarterly*, 15, 149-56.

——. (1984). Concepts of supervenience. *Philosophy and Phenomenological Research*, 45, 153-76.

——. (1989). The myth of nonreductive materialism. *Proceedings and Addresses of the American Philosophical Association*, 63: 31-47.

——. (1992). Multiple realization and the metaphysics of reduction. *Philosophy and Phenomenological Research*, 52, 1-26.

——. (1998). Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation. Cambridge, Mass.: MIT Press.

Lewis, D. (1966). An Argument for the Identity Theory, *Journal of Philosophy* 63, 17-25.

Lewis, D. (1972). Psychophysical and theoretical identifications. *Australasian Journal of Philosophy*, 50, 249-258.

Putnam, H. (1967). Psychological Predicates. In Capitan and Merrill (eds.), *Art, Mind, and Religion*. Pittsburgh: University of Pittsburgh Press.

Quine, W.V.O. (1969). *Ontological Relativity and Other Essays*. New York: Columbia University Press.

Smart, J.J.C. (1959). Sensations and Brain Processes. *Philosophical Review*, 68, 141-156.

What, If Anything, is Psychiatry?

DOMINIC MURPHY

Introduction

For much of the twentieth century American psychiatry was dominated by psychoanalytic approaches to mental illness[1]. These psychoanalytic approaches persisted even into the recent past, although within psychology there was always a strong behaviorist movement. But within the last thirty years or so another approach has superseded psychoanalysis as the dominant model in the study of psychopathology. This is the medical model, defined by one of its chief architects and theorists, Samuel Guze, as "using in psychiatry the intellectual traditions, basic concepts, and clinical as well as research strategies that have evolved in general medicine " (Guze 1992: 129).

The basic assumption of the medical model is that mental illness, like any other kind of illness, "represents the manifestations of disturbed function within a part of the body" (Guze 1992: 44). In mental illness the relevant body part is the brain. Guze is happy to countenance the existence of social forces and other forms of nonbiological causation in the production of mental illness but insists that the brain is "at the hub of psychiatric thinking" (Guze 1992: 59). The difference between Guze's medical model and the "biopsychosocial" account favored by some others (e.g. Wilson *et al.* 1995) is that while both approaches take into account a variety of social, psychological and neurological properties and causes, the biopsychosocial model is not hierarchical: it specifies no metaphysical or epistemic priority to any one element. The medical model, on the other hand, is explicitly brain-centered. It "specifies a hierarchy among its elements for conceptualizing and practicing psychiatry" (Guze 1992: 59) with the brain at the top. Later on, I shall look in more detail at this distinction between Guze's medical model and the biopsychosocial approach and provide some reasons for skepticism about Guze's account.

In this paper I am going to look at some of the issues surrounding the medical model, in an attempt to understand its rise, its current status, and its prospects. First I am going to provide some recent history, telling the story of the way in which issues of classification and diagnosis became central in psychiatry, since this is also the story of how the medical model rose to its current dominance. Second, I set out two ways of thinking about the medical model – the first is broadly sociological: we can think of the medical model, or "biological psychiatry" as a movement among American philosophers dissatisfied with psychoanalysis and especially interested in diagnosis and classification. The other way we can think about the medical model in psychiatry is to follow Guze's lead and to conceive of it as not just a name for a loose affiliation of individual scholars and therapists, but as a peculiar intellectual enterprise with a group of defining features, which mark it out as distinctive.

When we think of the new psychiatry in this way, I shall argue, a schism appears between some of the official anti-reductionist rhetoric of writers like Guze, and the way in which psychiatry is actually practiced, obsessed as it increasingly is with putative molecular and genetic explanations of mental illness. Another name for the medical model is "biological psychiatry", but this term, I will claim, is best seen as a way of doing psychiatry according to the medical model. Biological psychiatry stresses genetics and the influence of genes on neurochemistry. As advanced by Guze, the medical model is brain-centered; it makes room for other sorts of factors, but it is the brain that is at the center of psychiatric thinking. This appears to move psychiatry in the direction of neuroscience, but one of the striking things about the contemporary neuroscience of the mind is how cognitive it is, and this aspect of contemporary neuroscience seems squarely at odds with the reductionistic ethos of the molecular research strategies that currently dominate medically-informed work in psychiatry, despite what some of the adherents of the medical model seem to say.

The medical model, if it is to incorporate the brain into psychiatry in a plausible way, is going to need to be cognitive, and this is likely to collapse two distinctions: first, that between psychiatry and cognitive neuroscience, and second, that between biological psychiatry and the biopsychosocial approach to psychopathology. The first of these predictions is less controversial than the second, but it does the raise the question of what psychiatry will look like after the boundary between psychiatry and neuroscience is blurred. Traditionally, psychopathologies for which we develop a reliable biological etiology have simply disappeared from the province of psychiatry and become the property of general medicine. In part, the proponents of biological psychiatry are envisaging a repeat of the first part of this phenomenon of psychiatric displacement on a massive scale, since they foresee a future in which all psychiatric disorders become biological diseases. But they have not gone on to face squarely the question whether, if things do indeed come to pass in this way, psychiatry will have any future as an independent discipline.

Classification and the nature of mental illness:
the rise of the medical model.

Biological psychiatry, or psychiatry according to the medical model, can be seen as a reaction to the dominant psychoanalytic strain in the psychiatry which Americans were trained in before the 1970s. As well as a generally psychodynamic slant, that psychiatry was notable for a disregard of psychiatric classification, and the medical model began its rise thanks to the attention it paid to classification and careful diagnosis.

The first two editions of the American Psychiatric Association's *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association 1952, 1968: this manual is universally referred to as DSM-I, DSM-II etc. It is currently available in a revised version of its fourth edition) were based very heavily on psychoanalytic theory. Analysts tended to think that patients differed in degree rather than in kind, and taxonomy was thought of as having a substantial conventional element. The slighting of taxonomy in the psychoanalytic tradition showed up in a number of ways: the research was moribund, over 50% of patients went undiagnosed in one pilot scheme using DSM categories (Cooper 1998), and the great historian of psychiatry Henri Ellenberger airily suggested that systems of classification said more about the unconscious of the classifier than the nature of mental illness (Ellenberger 1963).

There was a sense then in which psychoanalysis did little more than pay lip service to classification, but this changed as the medical model jostled Freudianism out of the spotlight. As the psychoanalytic influence on psychopathology decreased, it was expected that new

bodies of theory would arise to put taxonomy on a sounder footing (Hempel 1963). Instead, the current classification aims to be "operationalized," "atheoretical" and "purely descriptive". It has not embraced a clear theoretical basis in biological psychiatry even though its proponents have usually come from that camp. The third and subsequent editions of DSM are supposed to be theory-neutral. Although the proponents of reform were often wedded to biological psychiatry the diagnoses themselves are not tied to any explicit general theory about underlying causes but are explicitly advertised as the product of "a descriptive approach that attempted to be neutral with respect to etiology." (DSM-IV-TR, p.xxvi). The reasons that the psychiatrists who masterminded DSM-III had for their embrace of a theory-neutral perspective were not just products of philosophical naivetÈ or wanton empiricism. It was quite reasonable to hope for a taxonomy that enabled practitioners from different backgrounds to talk to each other and it was also true that the etiologies of mañy conditions were (and remain) entirely unknown to psychiatrists, who lacked any comprehensive theoretical framework within which to think about the normal mind and its breakdowns. Whether this aversion to theory still makes sense in the light of subsequent developments in brain science is another matter, and one we will come back to. (It is worth noting that DSM-III did in fact explicitly exempt certain conditions from its strictures about theory-neutrality, to wit those whose etiology seemed reasonably well understood.)

The distrust of general theory among the architects of DSM-III was in part a result of their feeling that the psychiatry then dominant was not producing reliable knowledge about psychopathology. The remedy, they thought, was to apply to mental illness the standards and methods characteristic of medicine.

So what are the standards and methods of the medical model? The attention to diagnosis is one of them, as we have noted. It reflects the medical assumption that illnesses, including mental disorders, are distinct and identifiable units. Whereas psychoanalysis has tended to regard the symptoms of patients as different manifestations of a largely undifferentiated psychic crisis, medicine seems symptoms as tied to particular diseases with distinct natural histories – diseases, on this model, are discrete entities reflecting a disruption to a particular underlying function, with a characteristic and predictable course, outcome and response to treatment. In addition, they tend

to aggregate in families. This way of thinking about diseases marks a great shift in American psychiatry and accounts for much of the attention paid to classification during the last quarter-century, on the grounds that, as Goodwin and Guze put it, "classification in medicine is called 'diagnosis'" (Goodwin and Guze 1995: xi).

According to DSM-IV (xxi-xxii), a mental disorder is "a clinically significant behavioral or psychological syndrome or pattern" that is associated with distress, disability or a "significantly increased risk of suffering death, pain, disability or an important loss of freedom." But this must not just come about in a situation, such as a bereavement, in which psychological distress would be expected: it must be the expression of a physical, psychological or behavioral malfunction. A mental disorder, conceived by DSM-IV, is supposed to be the psychiatric analog of what would be called a disease by somatic medicine.[2] A disease is a set of symptoms that that are caused by a known destructive process. The range of destructive processes envisaged by DSM-IV is wider than that usually recognized by medicine; as we have seen, it includes "behavioral" problems.

There are over three hundred mental disorders recognized in DSM IV, but they are all supposed to meet four criteria, presented below, that together define a valid syndrome. A syndrome is not the same thing as a disease/disorder. A syndrome is simply a collection of symptoms that cluster together; it is a further question how the cluster may be explained. Moreover, very few of the DSM diagnoses are "validated" — the best way to think of validation is as establishing that a genuine syndrome exists, usually by discovering some physical "marker" which is highly correlated with the symptoms[2]. All DSM diagnoses are supposed to be valid, but, as I said, not many have actually been validated. In practice, "mental disorder' is used to refer to any condition recognized by DSM-IV (and some that aren't), whether or not the underlying cause is known, and whether or not the syndrome has been validated. In theory, a valid syndrome must meet the following four conditions, although sometimes one or more is fudged in practice:

1. *Natural History*. The Natural History of a syndrome is its characteristic clinical course and outcome. Kraepelin, for instance, distinguished the two major functional psychoses, schizophrenia and manic-depression, by noting that schizophrenia (which he called *dementia praecox* or "premature dementia") tends to be chronic and steadily and unremittingly worsens over a lifetime, whereas manic-

depressive illness (bipolar disorder) is episodic (Goodwin and Jamison, 1990). Some disorders are distinguished by the age at which patients tend to fall foul of them. Panic disorder, for example, tends to strike early in life, with a mean age of onset of 25 (Goodwin and Guze 1996), whereas Huntington's disease arrives in middle age.

2. *Family history.* Most mental disorders run in families, although there are apparent exceptions such as anorexia nervosa. Generally speaking, too, mental disorders seem to have a genetic component, with monozygotic twins usually more likely to share a diagnosis than fraternal twins.

3. *Differential response to treatment.* Two seemingly similar sets of symptoms can be distinguished if they respond to different kinds of treatment. When psychiatrists talk of differential response to treatment they usually have drugs in mind. More commonly, however, is one in which this "pharmacological dissection" does not work; most disorders respond to more than one drug. For example, both major depression and obsessive-compulsive disorder respond to Prozac.

4. *Laboratory findings.* Laboratory findings serve to link symptoms to underlying physical markers. But in many cases no distinctive physical markers can be reliably associated with a given symptomatology: correlations are often very low, and the markers also correlate with non-psychiatric problems. The following, for example, have all been claimed as indicators of schizophrenia: increased ventricle to brain-tissue ratio; abnormalities in the basal ganglia, corpus callosum or cerebral hemispheres; blood flow in the left globus pallidus; lowered platelet monoamine oxidase activity; hearing loss; eye tracking dysfunction; microscopic brain abnormalities; elevated levels of Borna virus antibodies; and morphological abnormalities, including high arched palates and low set ears (Goodwin and Guze 1996, McHugh and Slavney 1998). Laboratory findings associate a physical marker with a set of symptoms: they may not, on their own, explain why the symptoms are there.

So, current taxonomic practice, despite many rough edges, assumes that mental disorders are first, clusters of associated symptoms forming syndromes (picked out by criteria 1-4), and, second, expressions of malfunctions (and third, harmful in some way, but I will put that aside for now). Poland *et al.* (1994) call disorders that work like this "syndromes with unity": they claim that there are in fact very few syndromes with unity in psychopathology. Nonetheless, this picture of discrete disease entities, despite the criticisms that are often leveled against it, has become the dominant one in American psychiatry and is steadily gaining ground around the world. It should be seen as the key assumption of the medical model, since it marks the adoption of the perspective on disease that has long underpinned biological medicine.

Other aspects of the medical model

Along with this focus on distinguishing and validating identifiable illnesses, the medical model has adopted a number of other approaches that its proponents take to be definitive of medicine. Guze, for example, distinguishes three strategies of inquiry in medicine. The first is the epidemiological, which aims "to account for the incidence, prevalence and varied manifestations of illnesses in terms of environmental and demographic factors" (Guze 1992: 22). The second research strategy is the biological, which aims to understand "the clinical manifestations of disease in terms of the structure and function of the body at many different levels" (Guze 1992: 22). The epidemiological and biological strategies strive for generalizations across populations; they look for joints in nature into which classes of individuals can be placed. These first two strategies "come together at the clinical level" (1992: 22), where one employs the third strategy in examining and taking the history of a particular individual. It is Guze's contention that we can apply all these strategies in psychiatry as in the rest of medicine. Furthermore, insists Guze, these approaches enable one to use in psychiatry all the methods that any student of psychopathology might wish to employ – every possible causal factor, thinks Guze, can be fitted into the overall structure.

So far, my discussion of the medical model has centered on conceptual and diagnostic issues and been pitched at a fairly high level of abstraction. From this vantage point the medical model seems less like a clear-cut research program and more like a general orientation towards a set of issues. This may be true of any widely-held theory: David Hull (Hull 1988) has argued that it is folly to look for a set of beliefs held by members of an intellectual tradition (his example is Darwinism), advocating their construal as lineages of thinkers. In the case of the medical model in psychiatry, we may

not be dealing with a clear-cut set of shared commitments or a lineage in Hull's sense, but we can think in terms of an orientation away from traditional approaches and towards the picture of mental illness that I have outlined. (These "traditional approaches" include psychoanalysis and other psychodynamic approaches, but also behaviorism, Carl Rogers' humanistic therapy and sundry other theories.) In this sense we have, in the medical model, a way of thinking about psychopathology that makes up a prominent intellectual force that is steadily gaining ground within the wider culture. But it is consistent with an almost indefinite variety of particular applications in research, including some that are very different, even incompatible with each other. This way of understanding the medical model – as a broad orientation on psychopathology rather than a particular research program – gains in plausibility when we compare Guze's presentation to the self-image of theorists who cling to the biopsychosocial approach.

Guze's picture of the medical model is one in which broadly medical research strategies make room for cultural, familial and socioeconomic factors as well as biochemical ones. Proponents of the biopsychosocial approach also stress the interaction, in the causation and explanation of psychopathology, between biology, psychological factors (like learning history) and the social factors disclosed by epidemiological studies: when it comes to understanding the mentally ill, say adherents of the biopsychosocial model, "a single model of behavior is insufficient to permit full understanding of their conditions, unless it gives weight to all three systems (Wilson et al 1996)." Furthermore, advocates of the biopsychosocial approach are usually quite content to adopt the DSM classification, and the medicalized picture of disorders it envisions. That does not mean that biopsychosocially inclined clinicians and theorists do not have reservations about DSM. But reservations about DSM are widespread and are certainly held by many prominent advocates of the medical model. Goodwin and Guze (1995), for example, omit any mention of the personality disorders from their tour of major psychiatric syndromes on the grounds that the personality disorders are "too vaguely defined or too poorly studied to be useful diagnostic categories" (Goodwin and Guze: 317).

Wilson *et al* fault the medical model for neglecting nonbiological factors, treating all psychopathology as due to neurological and genetic abnormalities, and regarding therapy as exhausted by

medication. Guze regards this as a reductionist caricature of his view and insists that the medical model gives due weight to non-biological phenomena. But, we might wonder, if the medical model is so ecumenical, what is to distinguish it from the rival, "biopsychosocial" approach?

Guze dissents from the biopsychosocial model on the grounds that it is non-hierarchical, whereas the medical model in psychiatry follows the rest of medicine in recognizing the body and bodily processes as the essential factor. In psychiatry it is the brain that is the relevant part of the body, so in psychiatry according to the medical model it is the brain that is at the hub of psychiatric thinking – variations in brain function are "the heart of biological psychiatry" (Guze 1992: 59). However, Guze's metaphors of hub and heart are unhelpful. Also of little use is Guze's suggestion that the difference between the medical model and the biopsychosocial approach is that the former is hierarchical, with the brain at the top of the hierarchy. Does this mean that neurological explanations are assumed to be more important than other sorts of explanations? If so, the alleged unfairness of the claim that the medical model is reductionistic is hard to see. Does it mean merely that mental illness is always in some sense an expression of something going on in the brain, even when other factors are not of explanatory import? If so, then the difference with the biopsychosocial approach is hard to find, since proponents of the latter do not doubt that the brain is always implicated in psychopathology – to doubt that would be to give up on materialism.

It seems then that if the medical model is just an ecumenical materialism which stresses a view of mental disorders as discrete diagnostic categories then it is in fact no different from the biopsychosocial model. On the other hand, if it is the reductionistic enterprise its critics claim it to be, then it is clearly differentiated from other approaches to psychopathology, including the biopsychosocial, but this differentiation is bought at high cost. A reductionistic, molecular emphasis both renders the medical model less plausible as a general explanatory framework for psychopathology and moves it away from the neuroscience that it is supposedly inspired by, since that neuroscience is increasingly cognitive in orientation.

I have suggested that we can think of the medical model as a shared intellectual outlook uniting diverse approaches to

psychopathology. But there is also an avowed and conscious commitment to it on the part of people pursuing particular research agendas. If we want to understand how Guze's metaphors of hub and hierarchy translate into genuine theoretical and practical differences between the biopsychosocial and medical models we need to look at the medical model in action. I turn now to two recent attempts to outline a framework for psychiatry that draw on studies to exemplify their chosen point of view. These two papers illustrate perfectly the dilemma outlined above – either the medical model is no different from the biopsychosocial, or it is too reductionistic to be credible.

Kandel: biological psychiatry as reductionism

Guze thinks that biological psychiatry, which he says is an attempt to integrate psychiatry with the biological sciences, "is perhaps the inevitable outcome of following the medical model" (Guze 1992: 54). Bowing to this inevitability, in 1998 the distinguished neuroscientist Eric R. Kandel outlined a "biological framework" (Kandel 1998: 466) designed to unite psychiatry and neuroscience. Kandel's framework is based on five principles. The first is that all mental processes "derive from the operation of the brain" (Kandel 1998: 460). Kandel's second principle is that genes are "important determinants of the pattern of interconnections between neurons in the brain (Kandel 1998: 460). He goes on to conclude that genes thereby "exert a significant control over behavior" and that mental illness has a genetic component. Kandel's third principle is that gene expression is altered by learning, so that social or developmental factors also make important contributions to mental illness. Fourth, alterations in gene expression change neuronal connections and hence initiate and maintain behavioral abnormalities. Fifth, psychotherapy, in so far as it is effective, works by changing gene expression so as to alter neuroanatomy.

None of Kandel's five principles are inconsistent with the biopsychosocial approach, depending on how his language of gene expression is to be understood. His first principle is just a commitment to materialism that is shared by everybody working on psychopathology, included those working in the psychoanalytic tradition. Psychoanalysis was from the beginning conceived of as a materialistic theory (Sulloway 1979). Also, although the language

of genetic "control" is unfortunate, nobody with a biopsychosocial orientation would deny the claim that mental illness is in part genetic, nor that psychological and social factors affect behavior by altering the brain, either physiologically or anatomically. Indeed, the realization that brain structure and function can be changed by psychotherapy (e.g. Schwartz *et al* 1996) is often cited by proponents of the biopsychosocial model as a point in their favor since it indicates the capacity of psychological causes to have neurophysiological effects.

If we are to distinguish Kandel's version of the medical model from the biopsychosocial model, then much depends on what we are to make of his casting his discussion in terms of genes and their expression. One way of reading this is as tacitly assuming that, in the revolutionary new psychiatry Kandel looks forward to, we will ultimately explain everything at a genetic level. Although Kandel, as we have seen, countenances social and psychological factors as causes of alterations in gene expression, he says that "all of 'nurture' is ultimately expressed as 'nature'" (Kandel 1998: 460) and that "genes determine the phenotype". Kandel's language, then, suggests a picture in which we may acknowledge that a biological system is in a particular state because of a complicated array of causal factors, but that our explanation of the system's state occurs at only one level – that of molecular genetics. Kandel thinks that critics of his approach are uneasy because they think it is based on genetic determinism and that this is due to their being misinformed about the interplay of genetic and environmental factors. But, in fact, what worries critics of views like Kandel's is exactly what he stresses in his own account, the apparent faith that all explanations will occur at only one level of analysis, and that non-biological factors will be relegated to a subsidiary role. It is this interpretation of his five principles that separates Kandel's self-avowed biological framework from the biopsychosocial model. On this interpretation, what "biological psychiatry" comes down to as a particular implementation of the medical model is the search for reductionistic, typically molecular explanations of psychopathology in an intellectual structure in which other levels of explanation are neglected or given only a subsidiary role as causes of that which is to be explained in molecular terms.

Now, if what he means by "gene expression" is "behavior, neuroanatomy or whatever else seems relevant" then his picture is

one in which our mental life is instantiated in the brain but is caused and expressed at a variety of levels of explanation. And that picture is no different from the biopsychosocial approach. Again, we see that biological approaches to psychiatry either collapse into the biopsychosocial model in an attempt to make room for nonbiological explanations, or, if they try to remain biologically pure, they develop an implausible fixation on narrowly biological explanations.

Why implausible? Guze's and Kandel's view is that mental illness arises from underlying brain mechanisms that are impaired or malfunction in a way that is best explained by biology, even if caused by a complicated array of social or psychological factors. For example, the correlation between dopamine irregularities and delusional states or hallucinations is quite robust and widely accepted. Swerdlow and Koob (1987) propose explaining schizophrenia as overactivity in the dopamine systems of the forebrain, causing wide-ranging malfunctions in cortical cognitive and emotional processes. Whatever social forces have an effect on schizophrenics do not take pride of place in the explanations envisaged by advocates of the medical model. If Swerdlow and Koob (and other like-minded researchers) are correct, schizophrenia will ultimately be explained in terms of overactivity in dopamine systems.

On the other hand, not all psychopathologies are accompanied by discernible neurological differences. Although some physiological differences have been noted between anorexics and other people, for example, the conventional wisdom is that these are the result of starvation, and not themselves pathogenic (Goodwin and Guze 1996). The medical model does not seem to fit anorexia very well at all, whereas a sociocultural account looks most promising. It seems obvious, for example, that eating disorders are best understood as a consequence of the social pressures placed on young women who are told that desirability, or even virtue and worth, requires being thin (van Deth and Vandereycken 1995).

Pressure to be thin has increased recently (by 1978 the average Playboy centerfold weighed only 83 percent of the average female, compared to 91 percent in 1960 (Garner et al 1980)) but is quite long-standing: concerns among girls about their shape were recorded in the seventeenth century, fear of getting fat was seen in patients in the eighteenth and nineteenth centuries (drinking vinegar was one folk remedy) and medical journals discussed "morbid dread of fat" over a hundred years ago (Parry-Jones and Parry-Jones 1995).

Furthermore, we can trace the difference that exposure to modern western norms of female beauty makes; a study of girls at English-medium high schools in Pakistan found only one case of bulimia nervosa, but when the same investigators studied British schoolgirls from Indian and Pakistani immigrant families, they found a prevalence rate for bulimia of 3.4 percent. And studies show that young white women in America feel under greater pressure to be thin than do African-American females, and have correspondingly higher rates of eating disorders (Wilson *et al* 1995).

Accounts of mental illness stressing social factors seem to work well for eating disorders. Their epidemiology appears to track exposure to western conceptions of female beauty very closely, and we should look to the psychological effects of these cultural norms if we want to understand the immediate causes of eating disorders. Of course these psychological effects will occur in women's brains, but it seems reasonable to look to social and cultural inquiry for the best explanations of eating disorders. If we focus obsessively on abnormal brain activity, we simply obscure the actual etiology, since, as Miller puts it, if psychological trauma "fosters psychopathology via some mediating biological process, then the psychopathology is not fundamentally due to something biological. The trauma would be the fundamental factor, not the biological phenomenon that results from it, and we would still need to determine the mechanism by which trauma affects biology" (Miller 1996: 620).

To assert this is not to reject materialism or deny that genes may be involved, merely to recognize that our mental life can be described at many levels of explanation and that we cannot tell in advance the level at which the best explanation is to be found. Philosophers talk about levels of explanation (e.g. Lycan 1990) to convey the point that many complicated systems, including organisms, can be described in many different ways depending on what properties of the system we care about. An organism, for example, is made up of matter and hence has a description as an object that obeys physical laws. It also has a description in the language of chemistry, since chemical processes occur within an organism, and also in the languages of several biological disciplines at various levels from the molecular to the neuroethological.

Human beings also admit of description at several psychological levels; Marr (Marr 1982), for example, distinguished the computational and algorithmic levels of the visual system, and many

other theorists have argued that psychology also has an intentional level of explanation, in terms of beliefs, desires and other meaningful states. Nor do levels of explanation end with the individual. Human aggregations can be described at the economic level, the sociological level, and so on. One does not argue against the economic generalization that human beings seek to maximize utility by citing facts derived from DNA crystallography, and one does not make generalizations about psychological functions by pointing to lower-level facts, unless those lower-level facts are causes which have effects at the psychological level. The basic principle is, don't import facts from another level unless they have relevant effects at the level you are discussing. This point is clearly recognized in another attempt to outline an intellectual framework for a neurologically-informed psychiatry, due to Nancy Andreasen (Andreasen 1997).

Psychiatry as synthesis: Andreasen

Andreasen is struck by the convergence of complementary models of schizophrenia and depression deriving from a variety of disciplines, including neurology, the cognitive sciences and clinical psychiatry. In both cases, she argues, we can understand the conditions as cognitive pathologies instantiated in particular brain circuitry. This does not rule out, however, the study of mind and brain "as if they were separate entities" by "multiple and separate disciplines" using a "different language and methodology to study the same quiddity" (Andreasen 1997: 1586). But of course, as Andreasen's own discussion shows nicely, it is not just a matter of separate vocabularies and methods. Distinct disciplines attend to different properties of the mind/brain, and a fully adequate study of psychopathology must make room for all of them, plus others. As the example of eating disorders teaches us, we may need to draw on the resources of sociology or even cultural history to understand some disorders.

Andreasen conceives of her synthesis as a form of biological psychiatry, but in contrast to Kandell's molecular reductionism, she sees psychiatry as a cognitive discipline. Indeed, in her view it is one of the cognitive sciences: "the discipline within cognitive neuroscience that integrates information from all these related disciplines in order to develop models that explain the cognitive dysfunctions of psychiatric patients based on knowledge of normal mind/brain

function" (Andreasen 1997: 1586).

In the light of this straightforward assertion of pluralism it is hard to see what the difference is between the biopsychosocial approach and biological psychiatry. Andreasen's "scientific psychopathology" incorporates not just neuroscience but a variety of other disciplines. Does it put the brain at the hub of psychiatry, as Guze advocated? Yes, but in part because all the other disciplines Andreasen draws on study the brain in one way or another, even if they are not neurosciences. Guze's slogan might better have been phrased as: molecular neurology is the hub of psychiatry. But that, I have argued, is unduly reductionistic, and if we adopt Andreasen's more relaxed materialism the difference between the biopsychosocial and medical models disappears, and we are left with the attempt to relate clinical findings to a model of psychopathology that decomposes a cognitive model of a disorder into its component parts and then localizes those components in particular brain circuits.[3] Note that this process is driven by the cognitive description of the psychopathology – it is research on minds that guides research on brains, and historically it has always been so (Hatfield 2000).

In fact, this is just what we ought to expect if the medical model takes its own rhetoric seriously. Moving psychiatry closer to the biology of the brain just is to move it towards an increasingly cognitive discipline. Although there is much molecular biology being done on the brain, there is also an increasingly cognitive aspect to neuroscience, which includes the study of pathologies, and the old distinction between organic and functional psychopathology has collapsed. In the end, the medical model contains the seeds of its own annihilation by cognitive neuroscience.

Conclusion

So, what, if anything, is psychiatry after all? The most satisfactory recognition of what is happening to the subject may be summed up in Andreasen's claim that it is a field within cognitive neuroscience, and I have suggested that if the proponents of the medical model are serious about taking all the causes of psychopathology seriously then they will acknowledge that too. But taking Andreasen's perspective means making considerable changes in the education of psychiatrists and blurring the line between research in psychiatry and in the cognitive sciences. Ultimately, if psychiatry really is a

branch of medicine, but as such it must be the application of the knowledge we have of the working of a particular organ, and understanding the brain requires a variety of disciplines that look very unlike most medicine. If psychiatry is to be properly a branch of medicine, it needs to look less like medicine.

NOTES

1. I will restrict my remarks to American psychiatry: much of what is told here is true about psychiatry in other countries, though less so in some. France, where psychoanalysis remains influential, is one obvious exception, and phenomenological psychiatry is an important force in may parts of continental Europe. British psychiatry, on the other hands, has often led the way in the use of cognitive methods. However, the medical model is gaining ground all over the world (and forms of it have long been influential in other countries). As DSM (see text) becomes the standard diagnostic manual throughout the world (a process already under way, as acknowledged by at least one British textbook (Gelder, Gath and Mayou 1989)) the ethos of American psychiatry will grow with it.
2. Indeed, during the revision process which produced DSM-III the term "disease" was used, but then withdrawn when the American Psychological Association threatened to sue the American Psychiatric Association.
3. I have borrowed the idea of decomposition and localization from Bechtel and Richardson (1993): they argue that it is a basic strategy in many biological sciences to breakdown an abstract functional description of a system into components and then assign responsibility for carrying out those tasks to specific parts of a physical instantiation of the system.

REFERENCES

American Psychiatric Association (1968) *Diagnostic and Statistical Manual of Mental Disorders,* 1st edition. Washington, D.C: Author.

——. (1968) *Diagnostic and Statistical Manual of Mental Disorders,* 2nd edition. Washington, D.C: Author.

——. (1980) *Diagnostic and Statistical Manual of Mental Disorders,* 3rd edition. Washington, D.C: Author.

——. (1987) *Diagnostic and Statistical Manual of Mental Disorders,* 3rd edition, revised. Washington, D.C: Author.

——. (1994) *Diagnostic and Statistical Manual of Mental Disorders,* 4th edition. Washington, D.C: Author.

——. (2000) *Diagnostic and Statistical Manual of Mental Disorders,* 4th edition, textual revision. Washington, D.C: Author.

Andreasen, N. (1997). Linking Mind and Brain in the Study of Mental Illnesses: A Project for a Scientific Psychopathology. *Science,* 275: 1586-93.

Bechtel, W & R. C. Richardson (1993) *Discovering Complexity.* Princeton: Princeton University Press.

Ellenberger, H. (1963). Fallacies of Psychiatric Classification. In H. Ellenberger, *Beyond the Unconscious: essays of Henri F. Ellenberger in the history of psychiatry.* M.Micale (ed), New Haven: Yale University Press 1993: 309-327.

Garner, D. M., P. E. Garfinkel, D. Schwartz & M. Thompson (1980). Cultural Expectations of Thinness in Women. *Psychological Reports,* 47: 483-91.

Gelder, M., D. Gath & R. Mayou. (1989) *Oxford Textbook of Psychiatry,* 2nd ed. Oxford: Oxford University Press.

Goodwin, D. & S. Guze. (1995). *Psychiatric Diagnosis,* 5th ed. New York: Oxford University Press.

Goodwin, F. K. & K. R. Jamieson (1990). *Manic-Depressive Illness.*Oxford: Oxford University Press.

Guze, S.B. (1992) *Why Psychiatry is a Branch of Medicine.* New York: Oxford University Press.

Hatfield, G. (2000) Mental Functions As Constraints on Neurophysiology: biology and Psychology of Vision. In V. G. Hardcastle (ed) *Where Biology Meets Psychology: Philosophical Essays.* Cambridge, Mass: MIT Press: 251-71.

Hempel, C. (1965). Fundamentals of taxonomy. In C. Hempel, *Aspects of Scientific Explanation.* New York: The Free Press: 137-54.

Hull, D. (1988). *Science as a Process.* Chicago: University of Chicago Press.

Kandel, E. R. (1998) A New Intellectual Framework for Psychiatry. *American Journal of Psychiatry* 155 (4): 457-469.

Lycan, W. 1990. The Continuity of Levels of Nature. In W. Lycan (ed) *Mind and Cognition.* Oxford: Basil Blackwell: 77-96.

McHugh, P & R. Slavney (1998) *The Perspectives of Psychiatry,* 2nd ed. Baltimore: Johns Hopkins University Press.

Marr, D. (1982). *Vision.* San Francisco: W. H. Freeman

Miller, G.A (1996). How we think about cognition, emotion and biology in psychopathology. *Psychophysiology,* 33: 615-628.

Parry-Jones, W. Ll .& B. Parry-Jones (1995). Eating Disorders: Social Section. In G. Berrios & R. Porter (eds) *A History of Clinical Psychiatry.* New York: New York University Press, 602-11.

Poland, J., B. von Eckhardt & W. Spaulding. (1994) Problems with the DSM approach to psychiatric classification. In G. Graham & L. Stephens (eds), *Philosophical Psychopathology.* Cambridge, Mass.: MIT Press / A Bradford Book; 235-260

Schwartz, J.M., P. W. Stoessel, L. R. Baxter, K. M. Martin & M. E. Phelps (1996). .Systematic changes in cerebral glucose metabolic rate after successful treatment of obsessive-complusive disorder. *Archives of General Psychiatry,* 53: 109-113.

Sulloway, F. (1979) *Freud: Biologist of the Mind.* New York: Basic Books.

Swerdlow, N. & G. Koob (1987). Dopamine, schizophrenia, mania and depression: Toward a unified hypothesis of cortico-striato- pallido-thalamic function. *Behavioral and Brain Sciences,* 10: 197-245.

Van Deth, R. & W. Vandereycken (1995). Eating Disorders: Clinical Section. In G. Berrios & R. Porter (eds), *A History of Clinical Psychiatry.* New York: New York University Press, 593-601.

Wilson, G. T., P. Nathan, K. D. O'Leary & L.A. Clark (1996) *Abnormal Psychology.* Boston: Allyn & Bacon.

A Definition of Consciousness and a Framework for its Description in Physical Terms

TULSI DASS

Introduction

Consciousness is a term whose meaning we all take for granted in common parlance; yet it is difficult to give a formal definition of it in terms of well understood concepts. Although it is not essential to have a formal definition of consciousness in order to make some progress in understanding it, it is clearly desirable to have one. The author has constructed a definition of consciousness which satisfies some reasonable criteria. It is given in the next section.

The term 'stream of consciousness' was introduced by William James (James 1890); he described it as a process in which 'states of mind' succeed each other. One would like to have a concrete description of this process. By concrete I mean a description in terms of concepts that are well understood and which integrates well with the underlying neurophysiology. The simplest way to achieve this is to describe it as a physical process.

Such a description, however, poses new challenges. The main difficulty arises due to the subjectivity and irreducibility of the so-called raw feels, qualitative feels or qualia. A scientific description of pain, for example, would normally include its various objectively describable features like its location, intensity, behavioral effect, etc. This leaves out the subjective qualitative feel, the quale, of pain. It is not clear how to accommodate it in the traditional scientific framework. There appears to be a *prima facie* case for going to a broader framework (Edelman 1989; Chalmers 1996).

Neuroscientific studies of mental phenomena generally ignore these problems. They are usually concerned with exploring relationships between mental phenomena and processes at the neuronal level. One takes the view that the neurodynamic structure underlying

conscious experience would, with sufficiently advanced technology, permit detailed observations by external observers; the phenomena in question would then no longer remain exclusively subjective. No attempt is made to identify quantities that are directly related to conscious experience. This is analogous to trying to describe thermal phenomena (involving gases, say) directly in terms of molecular dynamics without introducing empirical observables like temperature (representing hotness) or pressure (representing the force per unit area exerted by the gas on the container walls).

The philosophical framework justifying such an approach goes by the name 'eliminative materialism' (Rosenthal 1971; Churchland 1981; Churchland and Churchland 1997). The justification is given in the context of intertheoretic reduction. A theory T_1 is said to be reduced to another theory T_2 if all the concepts of T_1 can be described in terms of T_2 and all the laws of T_1 can be derived from those of T_2 under suitable limiting conditions. Well known examples are T_1 = Newtonian mechanics, T_2 = relativistic mechanics and T_1 = thermodynamics, T_2 = statistical mechanics. In the context of mental phenomena, a fairly obvious choice for T_2 is (adequately improved) neuroscience. For T_1, the eliminative materialists take the so-called folk psychology (FP) which is an integrated body of generalizations like 'an angry man/animal is prone to violent behavior' reflecting commonsense understanding of mental phenomena. They argue, firstly, that FP is, indeed, a theory (and, therefore, assessable for its empirical adequacy) and, secondly, that it must undergo eliminative reduction (i.e. the concepts of FP must be simply discarded/eliminated) because (apart from its profound empirical inadequacy), it lacks 'active coherence' with modern science. In a recent paper with Manoj Srivastava (Dass and Srivastava 2000; a formalism has been evolved in which subjectivity is given a careful treatment and which does more justice to common sense notions of mental phenomena by refining FP into a *scientific* empirical theory. An improved version of it (especially in the treatment of qualia) along with some supplementary material is given in section III.

An appealing description of the stream of consciousness as a physical process appears when this framework is integrated with the physical picture of transient neuronal assemblies against the average (possibly oscillating) neuronal background. This is described in section IV.

Every physical process has associated with it a flow of information.

Some information theoretic aspects of neuronal processes (especially those relating to perception) have been discussed in the literature (see for example Churchland and Sejnowski 1992 and Rieke et al. 1997). Some interesting possibilities appear when information theoretic considerations are applied to OP. These are discussed in section V.

It is of some interest to have a second look at some of the works relating to consciousness in the light of OP. This is done in section VI. The last section contains some concluding remarks.

A Definition of Consciousness

I will give a few definitions or near-definitions given by some other authors before giving my own. First, an oft-quoted if-and-only-if kind of statement (and, therefore, potentially a definition) by T. Nagel (Nagel,1974):

> [A]n organism has conscious mental states if and only if there is something that it is like to *be* that organism — something it is like *for* that organism.

This conveys the essence of consciousness in an intuitively appealing manner; however, it cannot serve as a definition because ' the other side' of the if-and-only-if statement is by no means a well-understood concept.

Next, the definition of Searle (2000):

> Consciousness consists of inner, qualitative, subjective states and processes of sentience or awareness.

This definition may be called an enlightened common-sense definition; however, it suffers from the same inadequacy as the previous one: the term 'qualitative' is by no means well understood.

Among the various definitions and near-definitions of consciousness I have come across, I find the write-up for consciousness in *The International Dictionary of Psychology* (Sutherland 1989) most satisfying:

> Consciousness: The having of perceptions, thoughts and feelings; awareness. The term is impossible to define except in terms that are unintelligible without a grasp of what consciousness means. Many fall into the trap of confusing consciousness with self-consciousness — to be conscious it is only necessary to be aware of the external world.

I shall refer to the flaw mentioned in the second sentence in the quote above as circularity. This write-up has the essential ingredients of an acceptable definition but is itself not quite a definition as the author himself has implicitly admitted.

What are the criteria that a satisfactory definition of consciousness must satisfy? The following criteria almost suggest themselves:

(i) It should incorporate the commonsense understanding of consciousness.

(ii) It should employ only well understood concepts and should not be circular.

(iii) The definition should be able to serve as a concrete goal for the scientific explanation of consciousness. This means that the terms employed in the definition must have clear operational meaning from a scientific point of view.

(iv) It should preferably make allowance for the gradations of consciousness in various life forms.

Keeping this in view, I propose the following:

Definition of Consciousness: Consciousness is the property/faculty of some (possibly all) living beings incorporating the ability to have some or all of the following: perceptions, emotions, thoughts, willed action and a sense of self — with an awareness of having them.

(1) All terms employed have a clear meaning from the common man's as well as scientific point of view. In particular, awareness is a scientifically well defined concept and has a meaning quite independent of consciousness; inclusion of this term, therefore, does not imply circularity. The last part of the definition makes clear, for example, the difference between conscious and subconscious perception.

(2) The phrase 'sense of self' might appear problematical and a potential contributor to circularity; however, it can, at least provisionally, be taken to mean 'tendency of self-protection' which is operationally well defined. (Somehow the author likes this phrase to be an ingredient in the definition.)

(3) Some looseness in the definition has been kept deliberately. The property 'sense of self' should probably be an essential

ingredient of all forms of consciousness; however, being not quite sure, the author has chosen to remain non-committal on this.

(4) It should be obvious that the author has drawn substantially from Sutherland's write-up.

At this point one can proceed to define primary consciousness and higher order consciousness in its various gradations. It will, however, not be attempted here. For some enlightenment on this, the reader is referred to Edelman's book (Edelman 1989).

III. Framework for Describing the Stream of Consciousness as a Physical Process

A. The Philosophical framework

Let us recall the definitions of the terms "materialism" and "physicalism". (In the literature, definitions often vary from author to author. The definitions given below are the ones the present author likes.) Materialism is the doctrine that *all* phenomena in the universe can be explained in terms of matter and its dynamics. In its full scope, materialism makes allowance for laws governing matter which are outside the category of physical laws [i.e. laws expressed in terms of measurable quantities (or, more generally, quantities in terms of which measurable quantities can be constructed — Fermi fields in quantum field theory, for example) in the form of mathematical equations]. When these laws are restricted to physical laws, the restricted doctrine is called physicalism.

All successes of science so far (including those relating to genes etc.) have been in the physicalist framework. To accommodate consciousness, there appears to be, as mentioned in the introduction, a *prima-facie* case for broadening this framework. What is the proper framework ? To find an answer, we go back to some fundamentals.

B. The Primitive Elements of a Scientific Theory

Leaving aside the detailed hypotheses that have been employed in the construction of various scientific theories, the primitive elements of a scientific theory (which means, since consideration of fundamentals belongs to the domain of physics, primitive elements of a physical theory) may be identified as (Houtappel,Van Dam and Wigner 1965, Dass and Joglekar 2001)

(i) Observations/measurements (The ultimate concern of every scientific theory is the description and analysis of observations in some domain.)

(ii) Description of evolution of systems (This normally involves the concept of time.)

(iii) Conditional predictions about systems (given some information about a system, to make predictions/retrodictions about its behaviour).

The place where introduction of consciousness makes significant difference is item (i). The primitive entities relating to consciousness are experiences; to accommodate them we must replace 'observations' in item (i) above by 'observations and experiences': reduced to basics, the ultimate concern of a scientific theory general enough to permit treatment of consciousness, must be the description and analysis of observations and experiences. There is, in fact, a close relationship between observations and experiences and there is some redundancy in the phrase 'observations and experiences'. This redundancy, however, is harmless — at least in the present context.

C. The Problem of Subjectivity

As emphasized in the introduction, the problem of subjectivity should be faced and not bypassed.

What is the minimal requirement of objectivity for any theory (scientific or more general)? To answer this, we must have a close look at the rationale for the requirement of objectivity. The rationale is this: a theory is meaningful only if it is understandable unambiguously by an appropriate class of intelligent beings (supposedly equivalent in the relevant aspects of their intellectual equipment). It immediately follows from this that any theorizing (scientific or more general) of subjective matters is meaningful only to the extent that they can be objectivized (i.e. described in a manner so as to be understandable unambiguously by the relevant class of intelligent beings).

This appears somewhat disheartening. But there is hope. After all, in common parlance, we are able to communicate with each other about subjective qualitative feels like that of pain. (Otherwise we would not be bothering about constructing a theory for them!) To exploit this clue we must examine the role of and requirements on the scientific observers.

D. Scientific Observers

We intend to give a scientific description of consciousness in terms of observables. These are generally observer-dependent quantities.

Concrete mathematical descriptions of observables are given by scientific observers with respect to their reference frames. (A reference frame is essentially a choice of axes for fixing position coordinates and a clock for measuring time.)

Apart from being equipped with reference frames, traditional scientific observers are also (generally implicitly) assumed to be able to mutually communicate in a common language. *We augment this common language requirement by stipulating that the observers are adequately familiar with the folk psychological terminology with a clear mutual understanding of the meanings of various terms (arrived at through intersubjective communication).*

The doctrine of physicalism with this augmentation incorporated will be referred to as *Augmented Physicalism.*

E. Fundamental Postulate

The processes underlying all our observations and experiences are physical processes (to be described in the framework of augmented physicalism).

F. Observables Representing Qualia

Conscious experiences are generally composite. For example, one may be smelling a rose and also listening to some music simultaneously. We can decompose such conscious experiences into superpositions of elementary conscious experiences (which, by definition, cannot be so decomposed). To describe conscious experiences in terms of physical observables (to which we are committed), it is adequate to concentrate on the description of elementary conscious experiences.

It appears reasonable to assume that an elementary conscious experience (say, that of pain) can be analysed in terms of some conventional observables representing localization, intensity etc of pain and an irreducible qualitative feel (quale) of it. This quale, as understood by all observers through intersubjective communication, is the same (irreducible) entity; the observers can, therefore, assign to it a single mutually agreed upon symbol (say, 'P' for the qualitative feel of pain when some specific part of the body is pricked). The fundamental postulate then implies that this symbol must correspond to some observable. Since all observers have the same description for it, this observable must correspond to an observer-independent property of some neurodynamic subsystem. (The identification of these subsystems and their properties corresponding to qualia will be discussed in the next section.) These observables should be such

that it should not be possible to analyse them in terms of simpler observables which have direct correspondence with conscious experiences (irreducibility of qualia). We shall call them qualia observables. As noted by some authors (Lycan 1981, Churchland and Churchland 1997), what is irreducible (discriminationally basic) at a certain level of observation/experience, may well have a nontrivial structure at a deeper level. The qualia observables may well be complicated functions or functionals of neuronal or lower level (say, microtubule level) observables.

G. Objective Phenomenology (OP)

This is a thermodynamics-like empirical formalism in which mental states are described in terms of appropriate observables (just as, for example, the hotness of a gas is described in terms of an observable called temperature) called mental observables. (These include the qualia observables.) We denote the mental observables by M_a ($a = 1,2,\ldots,N$ where N is expected to be a large number). Recall that the pressure (P) and temperature (T) are expressible in terms of averages of molecular observables (Huang 1963; Reif 1998):

$$P = (m/3)\ (N/V)\langle v^2\rangle \quad T = (2/(3Nk))\langle E\rangle \quad \text{(1a,b)}$$

where m and v denote the molecular mass and velocity, V the volume of the gas, N the total number of molecules, E the total kinetic energy of the molecules and $\langle\ \rangle$ indicates the mean value. We expect similar relations expressing the mental observables in terms of (mean values of) functions of neuronal dynamical variables X_A ($A = 1,2,\ldots, N_1$ where N_1 is supposed to be a number much larger than N); we write these equations symbolically as

$$M_a = h_a\ (X_A). \qquad (2)$$

Equations (2) constitute the 'bridge laws' describing the reduction of OP to neuroscience.

Another important point to recall is that, for an ideal gas, the empirical observables V, P and T (although representing independent empirical concepts of extension, force per unit area on the walls and hotness) are experimentally observed to satisfy the equation of state

$$P\,V = n\,R\,T \qquad (3)$$

where n is the number of moles and R is the gas constant. This empirical relation is explained by appeal to molecular dynamics;

indeed, equation (3) follows from equations (1) by noting that <E> = N.(1/2)m $<v^2>$ and N k = n R. In general, a thermodynamic system has more than one equations of state (this number depends on the number of P,V type pairs for the system).We expect analogues of equations of state in OP:

$$G_p (M_a) = 0 \qquad p = 1,2, \ldots ,k, \text{ say} \qquad (4)$$

reflecting interdependence of mental observables. (For example, hatred, fear and aggression are related.) The relations (4) are expected to follow from the bridge laws (2) just as the equation of state (3) followed from the relations (1a,b).

It is to be noted that the bridge laws (2) provide a concrete realization of the mind-brain identity in the present framework. (In fact, the author arrived at the idea of this type of formalism by trying to obtain a formal expression of mind-brain identity in completely physical terms.) The usual objections to earlier versions of the mind-brain identity do not apply to these relations because these are relations connecting physical observables in a consistent scientific framework.

H. Conscious Mental States; the Stream of Consciousness

The space of all mental states [given by equation (4) just as the thermodynamic state space of a gas is given by equation (3)] involves cons-cious as well subconscious mental states. How do we characterize the conscious mental states? The qualia observables must play an important role here. ['Qualia are the souls of experiences' (Dennett 1997).] A typical qualia observable Q is expected to have a value Q_0 (typically zero) which represents the absence of the corresponding quale. We call it the 'dead value' of Q and all other possible values 'live values'. We call a mental state conscious if, in that state, at least one qualia observable has a live value; other states are subconscious.

Let P be the subspace consisting of conscious mental states; we shall call it 'phenomenal space'. At any instant *t*, the conscious mental state of the subject is represented by a point in P which moves as time changes; the curve so traced in P represents the stream of consciousness for the subject in question. The structure of the space P is expected to vary from species to species and, to a lesser extent, among members of the same species.

Global properties of P should represent some important features of the mental life of the subject. In particular, connectedness of P_1 expected to be related to the unity of self (Dass & Srivastava 2000).

IV. Connection of Objective Phenomenology with Transient Neuronal Assemblies

In the previous section, we introduced a thermodynamics-like empirical framework for describing mental states. We shall now try to make a closer contact with the underlying physics and relate the mental observables to concrete entities in neural dynamics.

Recall that, in thermodynamics, the variables V, P and T represent the properties of the gas as a whole. Similarly the mental variables are related to global (or semi-global) brain function. In this context, the objects which almost suggest themselves for consideration are the transient neuronal assemblies [Llinas and Pare (1996), Greenfield (1997)] whose role in cognitive functions of the brain has been steadily increasing in importance. These assemblies appear against an oscillating background (with the 40 Hz oscillations, etc.).

The situation here is, in its essential physical aspects, similar to the water surface with ripples in a pond; when a stone is thrown in the pond, wave-forms are generated whose shape, amplitude etc depend on the size, weight, momentum etc. of the stone. Here the ripples correspond to the oscillating background of neurons and the stone-generated wave-form to transient neuronal assemblies whose details correspond to the underlying sensations, thoughts, etc. An instantaneous mental state corresponds to a snapshot of the pond surface; the stream of consciousness corresponds to the changing scenario of ripples and the wave forms.

Now, in this example of the pond, the obvious way to proceed to describe the physical state of the water surface at any instant would be to introduce appropriate variables to describe the stone-generated wave-form and also some variables to describe the background with ripples. Similarly, it makes perfectly good sense to introduce quantities characterizing a neuronal assembly as a whole and try to analyze experiential data in terms of these variables and some other variables describing the oscillating background. We identify the totality of mental observables in OP with the assembly variables of all possible neuronal assemblies plus the variables describing the oscillating background.

What are the qualia observables in this picture? In this connection the author offers the following conjecture:

(i) An elementary conscious experience corresponds to the formation of a single transient neuronal assembly.

(ii) A quale signifies just the formation of a transient neuronal assembly. The qualia observable corresponding to it is just the formal observable identifying that assembly.

This conjecture is quite consistent with the known properties of qualia. Subjectivity of a quale corresponding to a neuronal assembly (say, W) reflects the fact that W belongs to the subject experiencing the quale (in contrast with, for example, an objectively observed piece of apparatus in a traditional scientific experiment). Moreover, we have an objectivized description of this subjective feature: a neuronal assembly is an objectively identifiable object; once the information about a quale experienced by a subject is communicated (say, verbally) to others, the corresponding neuronal assembly in the subject's brain can be observed by any observer. The irreducibility of the quale reflects the fact that it corresponds to W as a single entity. At a deeper level, we have, of course, a description of this entity in terms of the dynamical variables of individual neurons in the assembly.

The conjecture is experimentally verifiable. This can be done by exploring the correspondence between the qualitative feels and the transient neuronal assemblies.The conjecture, if valid, would:

(i) provide a partial justification for schemes which propose to deny or eliminate qualia (but otherwise take the transient neuronal assemblies properly into consideration); these schemes include, in particular, eliminative materialism mentioned earlier and the 'quining qualia' arguments of Dennett (Dennett, 1988).

(ii) imply that qualia are, at least in principle, realizable in the projects of artificial intelligence.

Even if the conjecture is found invalid, the formalism of sections III and IV has its value in giving a concrete example of how explanation of qualia can be reduced to explanation of structure and function thus proving invalid Chalmers's (Chalmers 1996) arguments to the contrary.

Information Theoretic Considerations

Information theoretic considerations are becoming increasingly important in the treatment of fundamental issues in science. Here we shall briefly consider how OP gets connected to the information theoretic approach in neuroscience and point out some useful

applications of information theoretic ideas in OP.

Information is the entity which makes the difference between knowing and not knowing; more concretely, it is a measure of the 'extent of knowing'. It serves to reduce uncertainty when more than one possibility for something exist. Information can be quantified. If an event can occur in two possible equally likely ways, then the message that one of them has occurred contains one bit of information; for n equally likely cases, the message that one of them has occurred has $\log_2 n$ bits of information. In general, assignment of probabilities $p_1, p_2, \ldots, p_n$ to a set of n possible outcomes has an information content given by

$$I = -\sum_{i=1}^{n} p_i \log_2 p_i = -[\log_2 p] \qquad (5)$$

where p is the random variable taking as values the n probabilities with the same probabilities. The quantity on the right in eq (5) is called the information entropy of the probability distribution. When the probabilities are conditional probabilities, the quantity is called the conditional information entropy. Information theory is the mathematical study of such quantitative measures of information and related matters.

In physics, information enters at the fundamental level when, to an observable (or, more generally, a set of observables) taking values in some range, either a value or a range of values is assigned, or, more generally, a probability distribution of observables is given. In a physical process, one has change of physical state of some system with time. A state determines either values of observables (in a deterministic theory like classical mechanics) or probability distributions of observables (in a probabilistic theory like quantum mechanics). It should now be easy to understand the statement that, every physical process has associated with it a flow of information.

Neuronal processes in the brain are physical processes and, therefore, have a flow of information associated with them. In particular, spiking of sensory neurons contains information (which is, of course, intuitively obvious). Information theoretic analysis of these processes (see, for example, Churchland and Sejnowski, 1992 and Rieke et al, 1997) throws light on several aspects of brain function.

Objective phenomenology employs some observables (the mental observables) which have appropriate ranges of values. As explained above, specification of mental state implies some information and

the stream of consciousness has associated with it a flow of information. Formally, the information theoretic aspects of the neuronal dynamics and that of OP can be related through the bridge laws (2).

An important use of information theory in statistical mechanics (see, for example, Robertson 1993) and elsewhere, has been the application of the maximal entropy principle to obtain thermodynamic laws for macroscopic systems. The point of interest for us is that the condition of maximization of entropy (which, as a physical quantity, is related to the information entropy by a multiplicative constant) serves to define a general direction for the multitude of microscopic processes in a macroscopic system. (In the thermodynamics of an isolated system, for example, it is the direction of approach to equilibrium.)

Now, there are some mental phenomena in which a choice of direction is involved. Two important examples are the phenomenon of 'attention' and operation of the so-called 'free will'. It appears plausible that each of these phenomena is governed by the condition of maximization of appropriate conditional entropy (of the probability distribution of the mental variables; the constraining condition is presumably defined by the external as well as internal (psychological) conditions under which the experiencing subject is operating.

A lot of work needs to be done before mathematical formalism along these lines is fully developed. Meanwhile, it is worthwhile trying to see if it is possible to make some useful predictions by making use of the formal expression of information entropy.

Some Earlier Approaches Reconsidered in the Light of Objective Phenomenology

In this section we shall have a 'second look' at some representative works in neurophilosophy in the second half of twentieth century. Only the essential points will be briefly dealt with. The order in which the authors are considered has been decided by the consideration of convenience of presentation. (The reader might find some 'logical sense' in it!)

(a) T. Nagel: In his very insightful and justly famous paper (Nagel 1974), Nagel pointed out the difficulty of accommodating the subjective aspect of conscious experiences in the framework of physicalism; to quote:

> If physicalism is to be defended, the phenomenological features must themselves be given a physical account. But, when we examine their subjective character, it seems that such a result is impossible. The reason is that every subjective phenomenon is essentially connected with a single point of view, and it seems inevitable that an objective , physical theory will abandon that point of view.

The response of the Churchlands (Churchland and Churchland 1997; this article represents, as far as the author is aware, their latest position) to the subjectivity problem is along the eliminativist line mentioned in the introduction. To quote:

> When the hidden neurophysiological structure of qualia (if there is any) gets revealed by unfolding research, then we will gain a new epistemic access to qualia, above and beyond each person's native and currently exclusive capacity for internal discrimination.

There is no attempt to describe qualia directly (i.e. without reference to the underlying neurophysiological structure). Our approach, described in section III, does more justice to the problem of subjectivity by making formal allowance for the use of folk theoretic terminology (supposedly 'standardized' or objectivized through intersubjective communication). This opens the way for introducing observables for qualia and other mental concepts.

Another point raised in Nagel's paper is that even if we manage to construct some working theory describing and explaining human conscious experiences, it will not be possible to do the same for other species — bats, for example. To quote:, 'Our own experience provides the basic material for our imagination, whose range is, therefore limited.'

The situation, however, is not all that hopeless. If the basic postulate stated in section III is valid, there is no reason why its generalization implying that 'the processes underlying the conscious experiences of all species are physical processes' should not be valid. One can, therefore, investigate those processes in greater and greater detail and construct more and more accurate theories for them.

(b) Chalmers: Chalmers's position, which he likes to describe as *naturalistic dualism* is explained in detail in his book (Chalmers 1996). This book has become famous for emphasizing the so-called *hard problem* of consciousness (which, as mentioned earlier, is the problem of explaining the subjective quality of conscious experience and the

irreducible qualia). Having noted that all success of science so far can be regarded as explaining structures and functions, he proceeds to say:

> When it comes to conscious experience, this sort of explanation fails. What makes the hard problem hard and almost unique is that it goes beyond problems about the performance of functions. To see this, note that, even when we have explained the performance of all the cognitive and behavioral functions in the vicinity of experience – perceptual discrimination, categorization, internal access, verbal report – there may still remain a further unanswered question: *why is the performance of these functions accompanied by experience?* A simple explanation of function leaves this question open.

A weakness is discernible in the argument given above (Church-land and Churchland, 1997): it implicitly *assumes* that subjective experience is something beyond what can possibly be described or explained in the scientific framework. This reflects at most a (fairly widespread) intuitive feeling and cannot be claimed to be based on a convincingly reasoned argument. In fact, we have seen in sections III and IV that the explanation of qualia, for example, can be reduced to explaining structure and function.

Having concluded that one must go beyond a physicalist framework for explaining conscious experience, Chalmers proceeds to construct a theory which has some psychophysical laws (laws connecting the mental and the physical domains) as its essential ingredients. As we have seen, there is no need to go outside physicalism; we shall, therefore not go into the details of these psychophysical laws.

I shall comment only on Chalmers's admittedly speculative account of the possible role of information as the basic entity relating the physical and the mental domains. He introduces the idea of realizing states of experience as elements of appropriate information spaces (phenomenally realized information) and proceeds to connect it to the traditional physical realization of information by introducing the concept of a *double aspect of information*; to quote:

> [W]henever we find an information space realized phenomenally, we find the same information space realized physically. And when an experience realizes an information state, the same information state is realized in the experience's physical substrate.

The trouble is that, whereas the physical realization of information can be made concrete in terms of values, or, more generally, probability distributions of observables, it is not clear how to bring about similar concreteness in the phenomenally realized information.

In OP, the information theoretic relation between the mental domain and the domain of the underlying neural dynamics follows directly from the bridge laws (2), as described in section V. In both the cases the information is realized physically; there is no need to introduce a mysterious double aspect principle. This appears to be the only possibility if one insists on a formalism which is predictive, verifiable and falsifiable.

(c) P.M. Churchland and P.S. Churchland: They are best known for their forceful advocacy of *eliminative materialism* (Churchland 1981; 1986, Churchland and Churchland 1997).We have dealt with it in the introduction, section III and in subsection (**a**) above.

(d) Mind-Brain Identity Theory: There are some variations of this theme, but the general idea of this theory is that mental states and processes are identical to physical states and processes in the brain, or, more generally, the central nervous system (Place 1956; Smart 1959; Rosenthal 1971; Davies 1995). Philosophical questions have been raised about the conceptual coherence of this identity (category error etc; see, for example, Shimony 1997). Our description of the stream of consciousness as a physical process can be thought of as a step further in the direction of this identity. Eq. (2) expresses this identity transparently. On the left we have mental observables which directly correspond with mental concepts and the right hand side represents neuronal dynamics in the brain. This realization of the mind- brain identity is in the same spirit as the one in which Eq.(1b) represents the identity of hotness with the molecular kinetic energy. This is facilitated by the introduction of an empirical observable T which represents hotness. Our mental observables analogously represent mental concepts. This program is, however, not complete because concrete expressions for the mental observables in the bridge laws (2) have yet to be obtained.

(e) Searle: Searle's latest position, which he likes to call *biological naturalism,* is described in his article (Searle 2000). Briefly stated, it is like this: Consciousness is entirely caused by neurobiological processes and is realized in brain structures. It differs from other

biological phenomena in that it has a subjective, or first person, ontology, but this subjective ontology does not prevent us from having an epistemically objective science of consciousness. The neurophysiological processes give rise to a unified field of basal consciousness (the term 'field of subjectivity' is also used by Searle for this) which is the normal state of wakefulness. Conscious experiences are perturbations or modifications of this field. As the source of this last idea, Searle quotes Llinas et al. (1998) according to whom consciousness is 'modulated rather than generated by the senses'.

These ideas appear intuitively appealing, but, unfortunately, lack concreteness. Searle does not offer any concrete description of the consciousness field, of its mechanism of generation and of its perturbation by the senses. He appears to leave open the possibility of this field being something more general than a physical field. If one insists on a formalism which is predictive, verifiable and falsifiable, then this field must be a physical field which must be describable in terms of appropriate mathematical variables. The formalism of OP offers a framework in which these ideas can be concretely realized. This can be done by using some of the mental variables to describe the consciousness field and some others to describe perturbations of this field by sensations.

The idea of employing some of the mental variables to describe the consciousness field can be appreciated better by noting that, in a body with non-uniform and non-stationary temperature, the temperature T will be a function of position and time and one can talk of a temperature field. Similarly, the formalism of OP makes allowance for some of the mental variables to be fields. The author hopes to say more on this theme in a future publication.

Conclusion

Keeping close to basics, it is shown that the proper way to deal with the subjectivity problem is to augment the traditional physicalist framework as discussed (i.e. to incorporate the folk psychology condition on the admissible scientific observers)

The reduction of folk psychology (FP) to neuroscience need not (and should not) be eliminative. The proper thing to do is to refine FP into a *scientific* empirical theory analogous to thermodynamics. Objective phenomenology (OP) is meant to serve precisely this

purpose.

In section III, whereas Eqs (1a,b) and (3) (belonging to thermodynamics) are concrete equations, the corresponding equations (2) and (4) belonging to OP are only formal. Finding concrete forms for these equations is a difficult task but worth undertaking to make concrete progress. The data on transient neuronal assemblies must be useful in this connection. Theoretical derivations of relations of the type (2) and (4), although quite challenging, appear to be within the domain of feasibility.

Thermodynamics played an important role in the development of statistical mechanics. Objective phenomenology can play a similar role (to a relatively limited extent though) for neuroscience. In particular, it should be helpful in framing the right questions for deeper studies/investigations in neuroscience.

REFERENCES

Chalmers, D.J. (1996). *The Conscious Mind: In Search of a Fundamental Theory*, New York: Oxford University Press.

Churchland, P.M. (1981). 'Eliminative Materialism and the Propositional Attitudes'. *Journal of Philosophy*, **78**, 67-90.

Churchland, P.M., and P.S. Churchland. (1997). *On the Contrary: Critical Essays.* (Cambridge, Mass.: MIT Press).

Churchland, P.S. (1986). *Neurophilosophy: Toward a Unified Science of the Mind-Brain*, (Cambridge, Mass.: MIT Press).

———. and T. Sejnowski (1992). *The Computational Brain.* (Cambridge, Mass.: MIT Press).

Dass, T., and Y. Joglekar. (2001).'Symmetries and Conservation Laws in Histories-based Theories'. *Annals of Physics (N.Y.)* **287**,191-228.

Dass, T., and M. Srivastava (2000). 'Thermodynamics-like Formalism for Mental States'. Unpublished research.

Davies, M. (1995). 'The Philosophy of Mind'. In *Philosophy: a Guide through the Subject*, ed. A.C. Grayling. (New York: Oxford University Press).

Dennett, D.C. (1988). 'Quining Qualia'. In *Consciousness in Contemporary Science*, ed. A. Marcel and E. Bisiach. (Oxford: Oxford University Press); reprinted in *The Nature Of Consciousness: Philosophical Debates*, ed. N Block, O. Flanagan and G. Güzeldere. (Cambridge, Mass.: MIT Press, 1997).

———. (1997). 'Qualia' in '*Conversations in the Cognitive Neurosciences*' ed. M.S. Gazzaniga. (Cambridge, Mass.: MIT Press).

Edelman, G.M. (1989). *The Remembered Present: a Biological Theory of Consciousness.* (New York: Basic Books).

Greenfield, S. (1997). 'Susan Greenfield talks to Stephen Jones at Tucson II', at the website www.culture.com.au/brain_proj/grnfield.htm.

Houtappel, R.M.F., H., Van Dam, and E.P. Wigner. (1965). 'The Conceptual Basis and Use of the Geometric Invariance Principles'. *Reviews of Modern Physics* **37**, 595-632.
Huang, K. (1963). *Statistical Mechanics.* (New York: Wiley).
James, W.T. (1890). *Principles of Psychology,* (New York: Holt, Rinehart and Winston).
Llinas, R. and P.S. Churchland, eds. (1996). *The Mind Brain Continuum,* (Cambridge, Mass.: MIT Press).
Llinas, R. and Pare. (1996). 'The Brain as a Closed System Modulated by the Senses'. In '*The Mind-Brain Continuum*' ed. Llinas and P.S. Churchland. Cambridge, Mass.: MIT Press.
Llinas, R., U. Ribary, D. Contreras, and C. Pedroarena. (1998). 'The Neural Basis for Consciousness'. *Phil. Trans. R. Soc. London Ser. B,* **353**, 1841-49.
Lycan, W.G. (1981). 'Form, Function and Feel'. *The Journal of Philosophy,* **78**, 24-50.
Nagel, T. (1974). 'What Is It Like to Be a Bat?'. *Philosophical Review,* **83**, 435-51.
Place, U.T. (1956). 'Is Consciousness a Brain Process?'. *British Journal of Psychology,* **47**, 44-50.
Presley, C.F., ed. 1967). *The Identity Theory of Mind.* (St. Lucia, Queensland: University of Queensland Press).
Reif, F. (1988). *Fundamentals of Statistical and Thermal Physics.* (New Delhi: McGraw-Hill).
Rieke, F. et al. (1997). *Spikes: Exploring the Neural Code.* (Cambridge, Mass.: MIT Press).
Rosenthal, D.M., ed. (1971). *Materialism and the Mind-Body Problem.* (Englewood Cliffs, NJ: Prentice-Hall).
Searle, J.R. (2000).'Consciousness'. *Annu. Rev. Neurosci.* **23**, 557-578.
Shimony, A. (1997). 'On Mentality, Quantum Mechanics and the Actualization of Potentialities'. In *The Large, The Small and The Human Mind,* ed. M. Longair. (Cambridge: Cambridge University Press).
Smart, J.C.C. (1959). 'Sensations and Brain Processes'. *Philosophical Review,* **68**, 141-156; reprinted in Rosenthal (1971).
Sutherland, N.S., ed. (1989). *The International Dictionary of Psychology.* (New York: Continuum).

Mind as the Third Level of Human Thought and Action

JAMES F. PERRY

This international conference on "Mind and Consciousness: Various Approaches" was called to create a better definition than those already in use for the words "mind" and "consciousness." There are many definitions of the word "mind" in our many special domains, but each of them in itself seems incomplete and together they are incoherent. So we gathered to create a better definition, a new paradigm. Not just any definition – anyone can invent a definition arbitrarily. The definition must meet certain standards. It must be useful for the work we aim to do. It must be relevant, understandable, and coherent. And I want to suggest that the concept of mind we agree on should serve to advance not only effective research but also effective human education and, not incidentally, peaceful human interaction throughout the world. Therefore I want to define "mind" and "consciousness" in terms of the logical distinction between *random, routine,* and *reflective* action. I urge that we define the mind according to its uniquely human function, which is, to choose reflectively the routine contexts or frameworks within which it usually operates. Everything in the world moves or changes randomly or routinely, but only humans can act reflectively. Only we can choose the contexts of our actions and take responsibility for the goals we seek and the purposes we serve.

I have traveled halfway around the world to be here today (at the conference) in the world's largest democracy and one of the world's oldest cultures because of two words: "*Buddha*" and "*Brahman.*" Both words originated in Sanskrit. The first meant "awakened" and the second meant "to cause to grow." What they may have come to mean is another matter entirely. These words and the ideas they were once used to identify, shorn of their later accretions of political and social signification, are in my judgment two of the most important ideas a person can have, and identify two of the essential

powers that differentiate a person from everything else there is. These two words name essential criteria for the presence of mind.

The three words I offer in return are simple, familiar words that taken together serve to describe a structure sufficient to locate mind and consciousness in the context of all possible human thought and action. Any human act, I suggest, is either (1) part of no plan at all, i.e., random; or else it is (2) part of some plan, i.e., routine; or else it is (3) about some plan, i.e., reflective. Think of these possibilities in terms of a game such as chess or cricket: we can do something that has nothing to do with the game, or else we can play the game, or else we can choose to play the game (or not). These three options are all there is. Everything we think or do occurs at exactly one of these three levels. These are the only possibilities. The words I use for these three levels of action are "random," "routine," and "reflective." Mind and consciousness, in this three-level structure, is reflective action, for the primary function of mind and consciousness is to render routines a matter of choice and responsibility. Mind is what Radhakrishnan calls "the unaffected spectator of the whole drama of ideas related to the changing moods of waking, dreaming, and sleeping." (Radhakrishnan, 1929: 32) It has only to be awakened and trained to its responsibility.

"But this is transparent!" you may say, and I agree. It is obvious. It is simple. It is quite clear. This is not a complex and sophisticated system. It's a simple structure containing just three levels, and everyone can count to three long before going to school. It is perhaps controversial, but it is more useful than any alternative system. Let us explore the model.

The first, the random level of functioning is called impulsive, whimsical, capricious, chaotic, arbitrary. It is unpredictable and normally unproductive (although there is a very special exception when creativity or discovery is the goal); but it is easy and exciting. It is the inevitable state of the untutored, but it is exhausting and usually self-destructive; and so those of us who survive long enough to procreate are those who have managed to transcend the random level of functioning to reach the routine level. Those who have not risen above the level of random functioning typically lead painful and confusing lives which they have no choice but to attribute to bad luck or malevolence of others.

We almost always "live, move, and have our being" in personal and professional routines that form the finite contexts of all our usual choices. The boundaries of our routines sort what usually matters from what usually does not matter. These routines of ours tell us what to expect in every ordinary situation. They are our homes, our sanctuaries from chaos, and the foundations of our plans and productivity. They are our familiar friends, our instruments, but they can easily become our gods, our prisons, our "*maya*" rather than our "*mukti*."

What sorts of familiar things count as routines? Consider the following list of examples: recipes, plans, games, institutions, trades and professions, contracts, ideologies, cultures, trails, scripts, jobs, careers, nations, habits, personalities, character, protocols, stereotypes, prayer as ritual, ceremonies, positions, default settings, study schedules, algorithms, budgets. Everything from the simplest gesture to the grandest civilization is a routine if it is a durable and predictable pattern and context of human activity.

Usually, the world over, routines are treated as absolute and unquestionable. They are taken as definitions not subject to falsification. They are, in a word, taken as divine. But suppose now that instead of defining our routine as valuable we take the empirical approach and observe its uses to find out whether in fact it is useful and for what purposes. Here are some of the uses we will find: promoting coherence across space and time, guiding expectations, coordinating diverse purposes, avoiding self-destructive behavior, communicating with others using a shared language, enhancing predictability to self and others, increasing identity and recognizability, managing surprise and stress, increasing productivity and long range planning, defending against persuasion and manipulation, promoting development of skills, judgment, and useful associations (such as perceptual, conceptual, affective, motor), enhancing our ability to recognize mistakes and other emergencies, to name several. This brief outline suggests why it might seem plausible to worship routines: they are valuable for so many reasons that it may seem unthinkable to consider the possibility of change. Routines will vary enormously, however, as to their breadth complexity, and stability, that is, as to the amount and quality of each function they provide, and also as to the amount of change in the person and the environment they can tolerate while still providing those functions. This fact creates the possibility of improving routines and comparing them with others. Hence the need for reflection.

Reflection is where the uniquely human level of mind enters the

arena. Reflection is the level at which routines themselves become objects of analysis and choice: *mukti.* Reflection is the activity of considering the nature and costs and benefits of a routine, relative to resources, purposes, and possible alternatives. At the reflective level we become capable of comparing routines with one another, and with the purposes they are adopted to achieve. At the reflective level, routines become means, options, and choices, rather than ends, definitions, and destinies.

Consider the function of such offices and institutions as legislatures, judges, arbitrators, diplomats, trustees, boards of directors. Their task is not simply to administer the law or rules as previously decided; a vital part of their task is to review those laws or rules to decide whether wisdom requires that an exception be made, or some amendment of the rule or law itself. We speak of conscience and prayer as review and analysis of routine, because we recognize that each person has the potential to perform this legislative function for itself, to learn to choose the very games we play, the goals we seek, the purposes we serve.

The functions of reflective action are varied: creation of routines and analysis of iatrogenic and nosocomial issues, revision and amendment of routines, comparison of routines in terms of costs and benefits (such as utility, elegance, learnability, coherence, efficiency, adaptability, fruitfulness), transition between routines (such as religious conversion, marriage, divorce, promotion, immigration, learning, retirement), correcting mistakes and resolving emergencies, coordinating and integrating diverse routines. All of these functions are vital from time to time if the purposes of the routine are to be served. Yet only a reflective person, awake and aware and adequately skilled at the reflective level of action, can perform these functions.

The idea of a routine, is central to this simple three-level system that all persons, men and women, of every age, class, and culture, are to share, as I believe they must. Leave the six-stage and eleven-stage and thirty-two stage systems for the doctoral candidates showing their dexterity. For the rest of us, and even for those learned doctors outside of working hours, three levels will be enough to explain the awakening and growing that is essential for the full becoming of the human mind. Leave the systems that discriminate between one group of human beings and another for those who value confusion, prejudice and conflict.

This conference aims explicitly "to evolve a common methodology and research goals across the disciplines for multidisciplinary research into mind and consciousness," and "to explore whether Indian theoretical traditions can offer useful insights in the field." I have traveled far to salute these goals and those traditions, but I suggest that there is very much more at stake here than simply increasing our capacity for scholarly cooperation. I suggest that what we need here is a coherent account of free will, which is a necessity for the future of democracy no less than the future of science and for the same reason, namely, to hear every voice and view all evidence.

Democracy requires understanding, for there can be no sustainable humane peace and cooperation between unequals in understanding. The twentieth century was the time when the human race at large at last began to awaken to the fact that not even a useful and beautiful routine is enough, without understanding, that is, without mindfulness. For those who are unequal in understanding lack shared understanding of the costs and the benefits of their choices. Those lacking understanding are more utensils than persons. They are puppets. The world over, those who are being damaged by the actions or policies of others may not be aware of the fact, and for that reason alone may not be able to act in their own defense either by their own acts or by calling for action by others.

A century ago when Darwin's theories of evolution were being hotly debated, a world-wide quest began to find pithecanthropus, the so-called "missing link" between the higher animals and civilized man. One wag suggested that we ourselves are that missing link. This is not an exaggeration. There is a vital logical criterion for human civilization that at present we do not meet. There is no common ground at the routine level, and never will be. At the routine level we are all different from one another, try as we might to achieve uniformity. At this early stage in the development of human civilization, we find that all over the world we have successfully risen above mere random impulse, and even above blind instinct. Yet we have hesitated on the plateau of routines to which we tend to devote ourselves and worship as divine definitions. The history of philosophy, it seems to me, is based on the crucial discovery that routine alone, however beautiful and however powerful, is not enough.

It is widely recognized that knowledge is power. So is under-

standing. Understanding is the greatest and most durable power, because understanding connects knowledge to the full range of purposes it might serve and moreover can correct itself. Historically, and for many reasons, elite populations having power were not inclined to share it with those they had no good reason to trust. Hence anything that would make understanding easy was rejected in favor of beliefs that promoted confusion and fear, whether of the supernatural or of the irrational. Recall the stated purpose of Plato's Myth of the Metals, which was to justify and sustain a hierarchic society by means of a carefully crafted myth. (Plato, *Republic*, 414f.)

This is a possible explanation for the scarcity of philosophy in the schooling of the masses. It is a possible explanation for the professionalization of philosophy, to contain it and direct it away from the general public. This may be why classic literature and especially modern literature about, for example, the distinction between mind and body, is so confusing. Perhaps it is intended to be. Since the best way to predict the future is to create it, by guaranteeing poor education we can predict that the poor will never understand, will never grow, will never awaken.

In some ways, however, this isn't confusing at all. At least since Aristotle it has been customary in western thought to distinguish head from hands, as in the colloquial distinction of "headquarters" and "hired hands." Accordingly we sort out management from labor, white-collar from blue-collar, and theoretical from applied science. In each case, of course, the former is the privileged part of the pairing. It is the difference between those who make policy, and those who implement it: persons make policy, utensils carry it out. C. P. Snow's 1958 book, *The Two Cultures and the Scientific Revolution*, focusing on the traditional distinction between Oxford's liberal studies and Cambridge's technical and scientific curriculum, stimulated the rise of today's superb technical universities such as MIT and our host institution, where there is are strong faculty and curricula in the humanities.

None of this argument is meant to belittle the importance and merit of routine. To the contrary: routine is as vital today as it ever was for survival and for progress. In premodern society, the shared task was to survive in a world over which people had little control. In addition to the many benefits of routine as listed earlier, there is also the fact that because of routine we are able to mean what we say, and to function together under the rule of law, not of man. In

this light it appears that to permit reflective thinking is to run the risk of making exceptions, which seems to replace the rule of law with the rule of arbitrary privilege, which is private law.

It need not be, and I think that for the sake of our survival and development into a fully functional and sustainable scientific and civilized community it must not be. It is vital not only for our long-term survival but also for our own moral worth that we seek to increase the capacity of everyone to assess evidence thoughtfully and to share in conversation about its meaning. It is vital for our shared humanity that everyone develop the capacity to be mindful and conscious, that is, to function reflectively. The foundation for understanding that I am proposing is transparent, which should be useful for this purpose. This is why I propose three levels of thought and action: random, routine, and reflective.

Let us review these three levels, starting with the most primitive, the random level. Any human act can be random, that is to say, part of no plan at all. Or else it can be routine, that is, part of some plan that endures for some time and typically includes a range of choices. Or it can be reflective, that is, it can be *about* some routine. (Please see the attached appendix for an outline of some instances, costs, and benefits of these three levels.) That is all there is. Any act must be of one and only one of these three types, logically located at exactly one of these three levels. To be sure, there are infinitely many possible routines, and perhaps even an unlimited number of kinds of routines, and moreover our routines are inevitably in flux owing to expanding awareness and a changing world. This doesn't matter. What matters is that routines are normally neither random nor reflective. They are tools.

Random action

Each of these levels appears abundantly in familiar experience as well as the literature of all cultures. Random action may be represented by the toss of a coin. Random action, which we call "whim," "caprice" or "impulse," and so forth, is on the one hand unpredictable, usually unproductive, and frequently self-destructive, albeit easy and exciting. We have only to note how much excitement means to most people and how much they will risk to obtain it to understand the appeal of random action. Those who simply react to impulse or other stimulus act randomly.

Routine action

In contrast to random action, routine action is predictable. Routine action is by far the most common form of action. That routine action is more common than random action should come as no surprise, given the self-destructive nature of random action. Those who act randomly do not survive. In addition to never achieving any goals except by accident, those who behave randomly tend to destroy themselves. They are indeed their own worst enemies. While those of us who are not privy to the plans other people are following may find that their behavior appears to be random, the fact that it is productive and not self-destructive tells us that what we are seeing is routine, not random: there is method in what appears to be madness.

Routine action is productive in countless ways, beginning with predictability. The human race has created many routines that each generation passes on to the next: social roles, trades, professions, academic disciplines, rituals, traditions, and many others. (Please see Appendix once more for a more extended list of examples.) Action once made habitual and skillful, i.e., routine, is relatively easy with manageable stress.

Reflective action

Machines function routinely, but something is missing for they do not choose their routines. Unlike humans, machines cannot be responsible for what they do, and the key to human responsibility is reflective action. Reflective action is the foundation of responsibility for routine. Any being that lacks the capacity for reflective action is a thing – an instrument, a tool, a utensil, a machine – rather than a person. Such a being lacks what thoughtful people have for millennia called a soul. Such a being is "*maya*," a prisoner within its fortress. It is not free, "*mukti*," in the largest sense and scope of the term. It cannot choose the game it plays and the purposes it serves, but can only play games and serve purposes chosen by others.

Now I want to place this concept of mind in the context of several familiar philosophical concepts. I am trying to persuade my reader that the distinction between routine and reflective action is the basis of the most useful distinction between non-mind and mind. To advance my cause I am going to offer you several examples from familiar experience to which this distinction can be applied, and

then I am going to survey very briefly a number of classic and recent texts. In both of these domains, familiar experience and classical literature, this distinction has been applied successfully. What I am proposing, then, is simply a codification and extension of a practice that has been found useful everywhere and always.

The analytic-synthetic distinction

The first of these familiar concepts is the analytic-synthetic distinction. The distinction between routine and reflective action can be elaborated in terms of the distinction between analytic and synthetic claims. Routines can be viewed two ways: as analytic claims, that is, as definitions, as ends in themselves or, instead, as instruments, that is, as factual claims about the way the world works and useful means to some further end, as idols or as utensils.

As idols they require no further justification, while as utensils they require evidence of effectiveness. (Note that it is one sort of activity, called "validation," to justify a move in a game, and a quite different and more elaborate sort of activity, called "vindication," to justify the choice of the game itself. Validation in terms of a coherent set of definitions is almost always easier than vindication in terms of an infinity of consequences mostly unknown.)

It is always possible to treat any human activity as a game, a pastime, an optional activity engaged in for whatever reason or for no reason at all but only by accident or habit. It is also possible to treat exactly the same game as an object of worship, a matter of fate or destiny or definition. Two classic Western languages, Greek and Latin, both have words identifying routine. In Greek, it is *ethos*; in Latin, *mores*. From these two words the English words ethics and morals are derived. But Greek and Latin also offer the roots of English words identifying the blind worship of routine: "zealot" and "fanatic." The very existence of these words tell us that the difference between the use of an artifact and its worship was familiar to those antique cultures, and as we have seen to still earlier cultures to the east, that is, here. The aim of human education must be to learn how to do both and to do them gracefully and wisely.

Cultures promote loyalty for many good reasons. Good habits and stable societies enable individuals to develop extraordinarily high levels of skill in valuable activities such as dentistry and aviation. As professionals we continue our training and not only maintain

but improve our skills because we live and work in an environment that protects us from such things as civil war, terrorism, pestilence, and starvation. The community as a whole benefits when its members are allowed to become extraordinarily productive. Productive citizens take better care of themselves and their fellow citizens (and of course they tend to pay more taxes) while at the same time being less of a drain on scarce resources such as police, courts, jails and hospitals.

In addition, conventional views of leadership hold that it is a mistake to reveal doubts or misgivings to those under one's authority. Such doubts and misgivings promote questions and reservations among the rank and file members of the community or institution and erode social stability. On this view it is vitally important to promote among ordinary people the view that their leaders have absolute authority based on absolute truth.

Historically, of course, this has never worked in the long run. From time immemorial those ordinary folk, having been marginalized in their own cultures, in time reject their leadership and create new cultures. The true task of culture is to help better cultures evolve within the context of the human community that already exists. Only when we think of our students and all younger generations as our successors and human equals will we feel comfortable helping them to understand this and participate in the conversation toward the future.

As soon as we all notice that routines, however complex, are artifacts, it becomes possible to see why they are useful: we built them that way. We constructed our routines, individually and collectively, because they appeared to serve, and up to a point did serve, our purposes. In general, routines increase predictability and productivity of many vital and useful kinds.

However, recognizing the contingent nature of routines introduces two other conditions to the situation. First, since we created our routines in finite amounts of time, with limited amounts of skill, vision, and information, it is not at all likely that our routines are perfect. Experience tells us that we can depend on them to contain an iatrogenic element: they will produce unintended and undesired consequences along with the results desired and intended. They will contain errors, mistakes, omissions, and contradictions from the beginning.

Second, we will find it wise to depend on the Second Law of Thermodynamics, or something like it, to describe how artifacts tend in the normal course of events to fray, erode, cool off, dissipate, deteriorate, and shatter, owing to corrosive and perverse forces either present internally or impinging from the environment, absent thoughtful infusions of energy, that is, maintenance. Just think of the energy that was put into organizing this conference. Did the conference run itself? Of course not. It took (and is as I write this still taking) constant effort. Nor does any career, or relationship, or any other institution or organization, survive without steady and thoughtful maintenance.

How can we expect people to trust in routine when we tell them that all routines are contingent and imperfect? Would it not be the wiser course to do as our ancestors did and assure the young that the routine was the work of a supernatural and perfect power?

This is not wise, and in the long run it will not work. In the first place, it is not necessary. Young people learn early that they grow, things change, toys break, people come and go. They can tell the difference between perfection and sufficiency. In the second place, the decision to promote supernatural (and thus untestable) agents begets several deeply pernicious albeit unintended consequences. It inspires others to do the same, and thus generates conflicts that cannot be resolved by any possible evidence. Moreover, it promotes scorn for evidence of any kind, especially evidence against the claimed perfection of the routine. In view of the myriad grounds for supposing that all routines must in time be inadequate, this has the effect of disabling people even for routine functioning as utensils.

The solution is to promote mindfulness, reflective thought, awakening, and to do this for all people. This will not only make them aware of their role in the process of developing better routines, it will also make them more appreciative of the achievements and character of their predecessors, themselves, and their fellow human beings. They will stop expecting perfection and will instead become proud of their progress.

Contradiction

The next philosophical concept to consider is contradiction. What is the unique use of mind here? Contradiction occurs as a result of random action, conflicts of routine actions, or careless reflection. Contradiction means, literally, two statements that deny each other, as in A and not-A. This is more than simple conflict. It is not, for

instance, simply being contrary. What is most worth noticing about contradiction is what follows from it. When we affirm and deny the same statement or proposition at the same time and in the same sense, the usual rules of deductive logic enable us to deduce all possible statements including the denial of all possible statements. The result of this is excommunication, confusion, and panic: language, thought, and action become impossible. Emotional commitments compound the crisis: we fall in love with our own beliefs, and when those beliefs are contradictory our judgment is heartbroken.

The medieval European logician John Buridan portrayed this condition in one of the most famous images in the history of philosophy: a hungry donkey, standing exactly midway between two succulent bales of hay, is unable to move not because he is chained or penned or threatened but because he is unable to choose. The animal, according to Buridan, starves to death in the midst of plenty, because it lacks a soul (a reflective self) and thus cannot exercise judgment and thus rise above its conflicting impulses! Such an animal is asleep, a robot, a zombie, a machine.

Such a state of being is literally diabolical. The word diabolical is generally taken to mean devilish but it is important to go back to the Greek roots of this word, *dia*, which means "two," and *ballein*, which means "to throw." Buridan's famous donkey was in a diabolical situation because he lacked the mind required to escape. He lacked the reflective level of being. The mind is therefore a sort of divine gift for overcoming contradiction.

Fiction

Consider now the mind in the context of the relations between an author, the characters in a story he has written, and the actors who might play those characters. Suppose I have written a story, perhaps for a movie to be made here in India. I as the author will know all I care to know about the characters, and will have complete power – in a word, author-ity – over them The characters, on the other hand, will have no such knowledge and power over me. This is because the characters occupy the routine level of being, their actions determined by the script which they have no power to change nor even to know about, while. I, the author, am at the reflective level of being relative to the characters in my story. I can descend from the reflective level of being to act upon the routine, to function as a

deus ex machina, stepping into the script to rewrite it. The characters cannot return the favor. Actors, of course, can function reflectively to transcend the characters they are playing, to make suggestions to the author, but the characters those actors play cannot even know that there are actors and an author. An individual's reflective self (Plato's "philosopher-king") can function as the author of that individual's routine self: we can write our own scripts. The mind is, literally, the ultimate authority over the self.

Plato's two levels of justice in the republic

In the *Republic*, Plato uses an astonishing example discussing the concept of justice. Socrates asks "whether any set of men – a state or an army or a band of robbers or thieves – who were acting together for some unjust purpose would be likely to succeed, if they were always trying to injure one another." Note the distinction Plato is making between acting successfully and acting for a just purpose. Both of these are required for true justice, according to Plato. At the routine level, all that is required for worth is effectiveness. But the act effectively done may be evil. Plato asks, "Take for example such actions as deliberating or taking charge and exercising control: is not the soul the only thing of which you can say that these are its proper and peculiar work?" The soul (or: mind) is a level of functioning, specifically, the level at which one chooses one's standards, goals, and purposes. Whatever does not do this well is unjust, and whatever does not do this at all is inanimate, that is, mindless. Hence there are four options, here ranked in descending order of worth: (1) to do a good thing and do it well; (2) to do a good thing and do it poorly; (3) to do a bad thing and do it poorly; and (4) to do a bad thing but do it well. The routine self may do well what the reflective self if conscious would not wish done at all. Without the mind, the best that we can be is excellent machines.

Descartes's two levels of being

In his first *Meditation*, Rene Descartes uses the familiar dream hypothesis, which holds that we cannot know now whether we will in future awaken, to demonstrate the possibility of transitions so fundamental as to seem identical to the process of awakening from a dream. In his second *Meditation*, he uses the less-familiar evil genius

hypothesis to reveal the astonishing presence of a level of self that maintains a continuous existence through any transition. This eternal self is the mind or consciousness. It exists through any routine transition, but it exists unnoticed until an individual makes the special transition from routine to reflective thought. This eternal self is the fundamental constituent of being for humans.

The dream hypothesis

At least since Plato in the west, and even earlier in the east, according to Radhakrishnan, the puzzling notion of awakening from a dream has attracted the attention of philosophers and poets. In the West this hypothesis is usually scorned, but the question it raises is devastating to all forms of arbitrary authority and especially to the simpler forms of empiricism ascendant there. The dream hypothesis asks: how do I know that I am not now dreaming and about to awaken to the ringing of an alarm clock the very existence of which is now unknown to me? The answer to this question is that I don't know. I have no way of knowing whether I am going to wake up and find out that everything important to me, and most especially the entire routine which I have occupied and used throughout the dream, has been but a dream, an illusion, a mistake. I don't know whether I am dreaming until after I wake up, and even then I don't know whether I will subsequently wake up once more. Nor does anyone, ever: that is the human situation. Life-changing transitions can occur to people who are already awake in the physiological sense of the word. We acknowledge this by using such terms as "born again," "rebirth," "renaissance," and "awakening" to describe them.

Three kinds of awakening

This description suggests that there are three kinds of awakening. The first kind of awakening is familiar to everyone who has undergone a life-changing transition from one routine to another – marriage, divorce, retirement, promotion, the sudden death of someone close and dear. Such a transition is categorically different from a change that takes place within a settled and enduring routine; it is a change of the routine itself, and calls, usually, for substantial and mindful changes of plan, priority, and habit. It may also call for regret and apology.

This experience of awakening is self-certifying in a way I now understand to have been described in Sankara's *advaita vedanta*. (Psychologists refer to the "aha!" phenomenon that occurs when a person recalls or realizes the right answer to a puzzle, the perfect solution to a problem. This truth is a matter of direct inspection, though such experience can be illusory.)

The second kind of awakening is the process of coming to the reflective realization that such transitions can happen again. This kind of awakening calls for the kinds of preparation that will help the person exercise his will wisely and well no matter what changes may occur.

The third kind of awakening, I suggest, following Radhakrishnan, is the process of realizing reflectively that almost no one is awake beyond the occasional spark. We are almost all sleepwalkers, automatons, robots, dreamers, "on autopilot," relative to what we might become.

The dream hypothesis is about the possibility that we might be mistaken about everything important, even our most fundamental beliefs about the world and our place in it. This possibility naturally is not often welcome to those in authority, and yet, as we will see in the next section, it is absolutely vital for scientific progress.

Given more time and space, we might survey the history of philosophy and psychology for further examples of the occurrence of these three levels. For instance, Aristotle distinguished the moral virtues from the intellectual virtues. Nietzsche distinguished slave and master morality. Sartre distinguished the *être en soi* from the *être pour soi*. In all three of these authors, the first reference is to the routine or habitual level of functioning, the second reference is to mind, awareness, responsibility. And the psychologist Lawrence Kohlberg called these three levels preconventional, conventional, and postconventional. I owe to him my understanding that these three levels are mutually exclusive and jointly exhaustive of all the possibilities.

Thomas Kuhn

Thomas Kuhn's famous *Structure of Scientific Revolutions* became one of the most popular works published in philosophy in the 20th century. In this work the author introduced the terms paradigm and anomaly to a wide audience in every arena of professional human

activity. Kuhn distinguishes the state of scientific product at any particular point in history from the process by which that product came to be. Kuhn describes every science as going through a process that starts with random experiments and observations, and in the course of time results in a shared, enduring model, which he calls the paradigm, of how that science is best learned and practiced. This shared paradigm then encounters anomalous experiences that do not fit the paradigm's categories and criteria, unpredicted and even unpredictable in terms of the paradigm. In time either the anomalies are successfully explained away as erroneous or irrelevant, or else the discipline (usually its younger practitioners) creates a new paradigm that suffices to explain and predict all the phenomena predicted and explained by the old paradigm, and also including the erstwhile anomalies. The paradigm, of course, is the routine. The reflective process notices anomalies and creates a new paradigm to explain both the old paradigm and the anomalies.

Kuhn's view of the evolution of knowledge rests crucially on the possibility of error. Scientists always assume there is something wrong with their data and their theories, and work ceaselessly to discover what it is. This is the price of progress, to admit that you might be wrong.

Amartya Sen

The very title of Sen's 1999 book, *Development as Freedom*, focuses on the process of growth. His view as an economist is that "development consists of the removal of various types of unfreedom that leave people with little choice and little opportunity of exercising their reasoned agency." (p. xii) This view, I think, squares with our experience. Choices we lack the power to make are choices we do not have, as are choices we do not know we have. Sen is not speaking of mere development in the sense of property or wealth, which can be stifling, but rather the development of human agency. His view resonates with the one I am presenting here. He claims that "as reflective creatures we have the ability to contemplate the lives of others.... It is not so much a matter of having exact rules about how precisely we ought to behave, as of recognizing the relevance of our shared humanity in making the choices we face." (Sen 1999: 283)

Our shared humanity is our mind, our creative potential. Just as we can be "of one mind" at the routine level about the beauty of Taj Mahal, so we can be, and, I think, are, "of one mind" in our reflective essence. We are in different places and times, with different accidents of birth, injury, and culture, fears, hopes, and memories; yet as humans we are one mind, one person, one being. To be aware of this is to recognize what has happened in the twentieth century. We have begun to learn that we cannot stop at the routine level of being, no matter how rewarding it may be in the short run. We must develop a shared awareness of our shared reflective being, our mind.

At this point I offer a summary comment from a recent book (Menand 2001: 399): Modernity is a condition a society reaches when life is no longer conceived as cyclical. In a premodern society, where the purpose of life is understood to be the reproduction of the customs and practices of the group, and where people are expected to follow the life path their parents followed, the ends of life are given at the beginning of life. People know what their life's task is, and they know when it has been completed. In modern societies, the reproduction of custom is no longer understood to be one of the chief purposes of existence, and the ends of life are not thought to be given; they are thought to be discovered or created. Individuals are not expected to follow the life paths of their parents, and the future of the society is not thought to be dictated entirely by its past. Modern societies do not simply repeat and extend themselves; they change in unforeseeable directions, and the individual's contribution to those changes is unspecifiable in advance.

Science and democracy

I am confident that we are going to find that brain activity no less than social activity fits this model. The true essence of mind is to act within a chosen context, e.g., to play a game, and at the same time to think about whether to stay there and play instead to stop or change it. It is this combination of capacities that makes human beings unique, and makes them at least responsible for their choices. It is this matching pair of capacities that makes it possible for human beings to grow without limit, to reinterpret and reconceptualize their beliefs. It is conscious awareness of the three levels of possibility for action, that makes it possible for every one of us in the human

community to practice our beliefs while understanding and appreciating the beliefs of others.

At every point in time, free will is going to be found to have its limits, but this does not matter. Like the pilings driven far enough into sand or marsh that will support a great building, our understanding and power need go no further than far enough to keep us learning, growing, awakening, to produce a harmonious life. Finite free will is yet free will. Our routines, which serve as our foundations for growing, need not be perfect (fortunately!) But if we are awake to the eternal possibility of awakening, then we may increase our power throughout our lives, and bequeath the result of our shared work to our successors.

Think about how parents teach their children that they have no choice but to become members (a ghastly metaphor: emigration as dismemberment!) of society. The world over, parents teach their children as best they know how, but the fact is that for the most part as Menand (following Alain Locke) said they simply use their children to serve their ancient ways. The children become permanently mindless, predictable and loyal but lacking the most important dimensions of free will. Such children lack what the western world has learned to call the soul. They are already damned before they die, and they die like sheep without having lived in the modern sense of the word. This may seem convenient for their parents, and indeed it probably is convenient, but it poisons the view of parent and child alike of what we humans are. Even the United Nations *Universal Declaration of Human Rights* mandates (Article 26, paragraph 3) that "Parents have a prior right to choose [that is, to limit] the kind of education that shall be given to their children." So long as this is the case, children, male and female alike, will be condemned to lives of mere routine and world peace even as an ideal will remain mere fantasy.

To succeed in teaching the reflective level of functioning, a shared vocabulary will be necessary. David Perkins recommends what he calls a "metacurriculum," that is, a vocabulary about the vocabulary of the regular (routine) curriculum. To share this vocabulary with our students is to make them members of our own linguistic community. In the words of Frank Smith, it is to treat them as prospective members of our club rather than merely as utensils to be used by the members for their own benefit.

The alternative is to teach children that they have no choice but to accept the routine, whatever it may be. It is to teach children that we, ourselves, have no choice. That is to say, it is to teach that there is no such thing as mind. But this inevitably generates a paradox. Reflective children will respond thoughtfully as follows. You say that there is no choice. So you are saying that the reason you are saying this is not that it is true or that you have grounds for believing it to be true, but rather you have no choice but to say this. You are saying that you are a robot, and I must ask: why should I obey a robot?

There is a fundamental sense in which we are not conscious if we do not know that what we are doing is optional for those doing the same thing consciously. If we do not know that we can stop, or that there is an alternative choice available, we are acting under compulsion. Without descending into the complex theories of Freud and Jung, I think it reasonable to consider simple and familiar instances, e.g., where a traveler does not know that there is a later train and so makes absurd and even dangerous efforts to catch the early one. Such occasions lead people to say, "I had no choice," or "I didn't know what else to do."

As long as people say such things about the routine contexts in which they were raised, so long will we have disagreement within every profession and discipline, and war between cultures. From a pragmatic point of view, all this conflict may seem to serve some sort of evolutionary aim, testing the true power of an academic clique or a nation. But this is an illusion, for the result of conflict is to obscure the accuracy and completeness of a view behind the veil of the accidental and arbitrary power of those who hold it.

The quest for a coherent account of mind and consciousness is a quest for an obvious criterion of humanity. It will not suffice to construct elaborate schemes of levels far too complex even for many practitioners to use effectively. The structure must be simple and obvious: it must be, in a word, transparent. Such a foundation will offer a framework that almost everyone (other than infants and the institutionalized) can share, offering predictability and control to all, and thus, responsibility and a full measure of functioning equality. This is what is at stake here. I hope that what I have offered here will be of some use to my colleagues and fellow human beings, to our shared community the world over and to our shared posterity.

BIBLIOGRAPHY

Krishnamurti, J. (1973), *The Awakening of Intelligence* (New York: Harper Collins)

Kuhn, T. (1962), *The Structure of Scientific Revolutions* (Chicago: Univ. of Chicago Press)

Menand, L. (2001), *The Metaphysical Club* (New York: Farrar Strauss Giroux)

Perkins, D. (1992), *Smart Schools: From Training Memories to Educating Minds* (Glencoe, IL: The Free Press)

Perry, J. (1975), "Plato's Dream Hypothesis: A Meta-Communicative Statement," *Man and World*

——. (1998), "The Dream Hypothesis, Transitions, and the Very Idea of Humanity," *Proceedings of the XXth World Congress of Philosophy*, Boston, USA, 1998 URL http://www.bu.edu/wcp/Papers/Teac/TeacPerr.htm

Radhakrishnan, S. (1929), *Indian Philosophy*, 2nd edition (London)

——. (1936) *Foundations of Civilization: Ideas and Ideals* (Delhi: Orient Paperbacks)

——. (2000) *Search for Truth* (Delhi: Hind Pocket Books)

Sen, A. (1999), *Development As Freedom* (New York, USA: Anchor Books)

Smith, F. (1990), *To Think* (New York, USA: Teachers College Press)

——. (1986) *Insult to Intelligence: The Bureaucratic Invasion of Our Classrooms* (Portsmouth, NH, USA: Heinemann)

APPENDIX

Random, Routine, and Reflective

Random (Preconventional, unpredictable, unproductive, self-destructive)

Examples: "Brownian" motion, caprice, coin tosses, whimsy, impulse, two (or more) masters, recipes, or other routines

Functions: spontaneity, distraction, exploration and discovery, novelty and variety, surprise and excitement, disguise, deception, confusion

Routine (Conventional, predictable, productive)

Examples: institutions, recipes, trades and professions, contracts, games, ideologies, cultures, trails, scripts, jobs, careers, nations, habits, personalities, character, protocols, stereotypes, prayer as routine ritual, ceremony, plans, positions, default settings, study schedules, algorithms, budgets

Functions: coherence across space and time, guiding expectations, coordinating diverse purposes, avoiding being "your own worst enemy," communicating with a shared language, predictability to self and others, identity and recognizability, managing surprise and stress, productivity and long range planning, defense against persuasion and manipulation, development of skills and judgment, development of useful associations (such as perceptual, conceptual, affective, motor), recognizing mistakes and other emergencies.

Reflective (Post-conventional, conscious of and responsible for routines, creative)

Examples: legislatures, judges, arbitrators, diplomats, trustees, boards of directors, amendments, conscience, consciousness, mind, soul, prayer as review and analysis of routine

Functions: creation of routines and analysis of iatrogenic and nosocomial issues,

revision and amendment of routines, comparison of routines in terms of costs and benefits (such as utility, elegance, learnability, coherence, efficiency, adaptability, fruitfulness), transition between routines (such as religious conversion, marriage, divorce, promotion, immigration, learning, retirement), correcting mistakes and resolving emergencies, coordinating and integrating diverse routines

Combinatorial Perspectives

Varieties of Supervenience

SANJAY CHANDRASEKHARAN

Supervenience is a concept borrowed from moral philosophy by philosophy of mind to solve the mind-body problem. The mind-body problem is actually a set of problems, among which the major ones are the Consciousness problem (Illustrative instance: how can physical states have *felt* qualities?) and the Intentionality Problem (Illustrative instance: How can physical states be *about* something, or be *directed* towards something else?). One of the major physicalist strategies for solving the mind-body problem has been *reduction*, i.e., the strategy is to argue that the mind is "nothing-but" some physical state. However, many physicalists have found this route very unsatisfactory, so there are also non-reductive physicalist theories of the mind (theories which hold that the physical level is the fundamental level of understanding, but which also hold that the mind cannot be *reduced* to the brain or any other physical level). Most computational theories of the mind (See for instance Newell's SOAR model, Anderson's ACT and ACT* models, and works by Chomsky, Minsky, Fodor and Pylyshyn) belong to this category. Supervenience is used by this kind of physicalist theory to explain the mind-body relation.

The structure of the paper is as follows: I first outline two philosophical notions, supervenience and multiple realizability. Then I classify supervenience into two types and describe them. I show that the first type does not solve the mind-body problem. Then I argue that the second, although useful in solving the mind-body problem, is in conflict with the notion of multiple realizability (the thesis that the mind can be instantiated by things other than the brain, which is the philosophical basis of all Artificial Intelligence (AI) efforts). Finally, I propose that through the Indian concept of *swasamvedana,* the conflict between supervenience and multiple realizability could be resolved.

Supervenience

Though Davidson introduced supervenience to philosophy of mind, the notion of supervenience considered here is the one laid out by Kim (1993). Here is a non- rigorous definition of supervenience, based on Kim (1993): *If A supervenes on B, any change in A is a change in B, but not vice versa,* where A and B can be objects, events, or state of affairs.

It may be applied to a strictly internalist story about the mind as follows: any change in mental state A is a change in the brain state B, but all changes in brain states are not changes in mental states. Let's call this supervenience-I, for supervenience-Internalist. This is the first kind of supervenience and I shall be arguing against this kind of supervenience.

The formal notion of supervenience, as outlined by Kim, has three versions, the weak, strong and the global. Here, I outline only the weak and strong versions below, because Kim equates the global with the strong one.

Weak supervenience: A weakly supervenes on B if and only if necessarily for any property F in A, if an object x has F, then there exists a property G in B such that x has G, and if any y has G, it has F.

Strong supervenience: A strongly supervenes on B just in case, necessarily, for each x and each property F in A, if x has F, then there is a property G in B such that x has G, and necessarily if any y has G, it has F.

Multiple realizability

The notion of multiple realizability I am using here is the garden variety one; so common that it is taken from the World Wide Web. Multiple realizability is defined as follows (from the online dictionary of Philosophy of Mind):

> If two events can be tokens of the same mental type (e.g., they are both thoughts about Vienna), while being tokens of two different physical types (e.g., one is a pattern of neural activations in a mammal's brain, one is a certain distribution of electrical currents in a creature made out of silicon), then the mental type in question is multiply realizable.

Less rigorously, the multiple realizability thesis says that the mind is not dependent on the brain as a physical system, and creatures made of metal, plastic or blue cheese can have a mind.

Two kinds of supervenience

The discussion of supervenience is largely done in the abstract. I would like to bring it down to a more concrete level using examples. One reason for doing this is to show that there are different kinds of superveniences, and therefore, appeals to supervenience have to specify which of these superveniences they are appealing to. Since this requires picking and choosing between superveniences, it shows that the supervenience relationship is a norm that is applied to the mind-body problem and is not a brute fact.

First, a bit of clearing of philosophical ground: I only consider the supervenience of mental content here; the issue of consciousness is not considered. Second, I don't consider the notion of supervenience of explanations, which considers one narrative (say of the mind) supervening on another narrative (say of the brain). Three, my examples are positioned against *internalist* models of the mind. These are models which hold that all mental states are *just* brain states. Finally, my sympathies are with an interactionist view of the mind, where mental states supervene on a brain-environment complex, instead of just the brain.

The supervenience relation, as seen in the world, can be categorised broadly into two types. Let's call them Conventional Supervenience and Physical Supervenience.

1. Conventional supervenience

Roughly put, this supervenience relation needs observers — A supervenes on B by the setting up of a convention by a group of people. A good example for this kind of supervenience is the dollar bill, usually put forward as an example of the token-identity theory, which is closely associated with functionalism and multiple realizability. This is the kind of supervenience that naturally supports multiple realizability and AI.

According to token-identity theory, each mental state is identical to a particular physical state (it is nothing more than that state), but there is no identity between types of mental states and types of

physical states. The value of the dollar note in your pocket could be constituted by a piece of paper (a one dollar bill) or a piece of metal (a coin). Note the following points.

- The bill and the coin belong to the same monetary type.
- The bill and the coin do not belong to the same physical type.
- Hence there is no identity of type between the monetary (being a dollar, having a value) and the physical (paper, metal).

The supervenience of software on a physical machine is similar to this. Notice that the above description does not say *how* there can be something like a monetary property, or the physical mechanism of the monetary property.

In fact, both the dollar bill and software presuppose a set of observers for whom the supervenient property (being a dollar bill and being, say, ASCI characters) makes sense. The function of the dollar bill and the function of the ASCI character set are established by convention, not by nature. Once this function is established, by a convention which says that this function supervenes on a physical object or state, the same property can be realized by different physical objects or states (by coins in the dollar case and by different kinds of machines in the ASCI case).

This kind of supervenience is not a valid one to explain the mind-body problem, because it presupposes an observer, or a set of observers. For the mind to be like a dollar bill, there has to be a convention by which people agree on psychological states supervening on physical states.

Objection

For there to be a convention that property A supervenes on property B, there need not be an observer. The convention can be established by a meaning-giving system. The notion of a meaning-giving system can be illustrated using a counterfeit coin. What makes a coin counterfeit is the fact that it does not originate from the government mint. And the government mint is not an observer, it is a "face-less" entity.

To treat this objection, let me first outline a real-life version of multiple realizability. In the early 1980s, India had a problem with producing coins. So there was a shortage of coins in circulation, and bus conductors in Delhi had a hard time giving people change. The conductors hit on the following solution: instead of returning change, they issued passengers blank tickets of smaller denominations. These tickets could be used for travel in other buses, or turned in at a local transport office to collect money. Now comes the interesting part. Shopkeepers started accepting these tickets as currency. So you could buy cigarettes with bus tickets!

Notice that a bus ticket not issued by a conductor will be considered as counterfeit. This is an instance of multiple realizability of value—a kind of twisted version of the dollar/coin example. But the twist is interesting.

What the example shows is that the "value-carrier" (coin or bus ticket) is not dependent on a "face-less" meaning-giving system. It is dependent on a group of people agreeing to an instrument of transfer. So the observers are back, and therefore conventional supervenience is not good enough for solving the mind-body problem.

There is more to the example, though. Notice that in the bus ticket example a *full transfer* of value (content) takes place from one physical base to another. So the question is: If the bus ticket is a bonafide case of supervenience created by a "face-less" entity, how does the transfer of content take place? How can a bus ticket suddenly become currency without its physical base changing? Projecting to the mind-body problem, this implies that my mental state can suddenly change into something else without any change whatsoever in my physical state. So supervenience-I doesn't hold.

However, as we shall see, the conventional supervenience relation need not be thrown away, it can be used in a way to support the multiple realizability thesis. Let's look at a more plausible supervenience candidate now.

2. Physical Supervenience

There is no convention associated with this kind of supervenience. It is a naturally occurring physical relation, something that exists independent of observers. In this relation, A supervenes on B *by virtue of B's physical properties.* However, as we will see, not all physical supervenience relations strictly follow the definition supervenience-I needs. I present two examples below that do not meet the requirements of supervenience-I. The examples are presented as analogies as well as illustrations. I encourage readers to think of the mind-body relation in analogy while considering them.

Physical Supervenience I: Consider magnetism. It is a natural supervenience relation. The property of being a magnet (the magnetic force) is supervenient on a physical substance. Notice two things, though.

1) A change in the magnetic force *need not come from* a change in the underlying physical substance. It can be caused by electric currents, by the presence of another magnet, or the change in the magnetic poles of the earth. In other words, if we consider the base B as the magnetic material, the supervenient property A can change without a change in the properties of the base.

2) The magnetic force cannot supervene on all substances. It is multiply realizable, but there are restrictions on what materials can realize it.

The same story can be told about electric charges. Magnetism could be made to fall in line with Kim's definition by enlarging the base property B to include everything in the environment that changes property A, including electric currents, other magnets, and the poles of the earth. But this would make supervenience a non-local relationship and not useful for the internalist.

In the case of the mind-body problem, a non-local supervenience relationship implies that mental content supervenes *physically* on the brain plus the environment. This conclusion is strengthened by the following example.

Physical Supervenience II: Consider an image reflected on a mirror. This is another naturally occurring supervenience relationship—the property of being an image supervenes on a physical substance, namely the mirror.

However, it is hard to tease out what exactly the image is supervenient on. The image is dependent on a number of factors – nature of object reflected, position of object, the position of light, the nature of the light, and the nature of the mirror. A change in any of these will change the image. Notice the following:

1) If we consider the mirror as the physical base, a change in the image is *not* a change in the underlying base.

2) If we consider the whole system of mirror, object and light as the base, then the image is supervenient on a complex of things, involving objects, properties (of being a mirror and being an object that reflects light), a fundamental force (light), and a particular configuration of the objects, properties and the force involved. This means supervenience is not a local relationship.

3) Given 2, it is an open question whether the *same* image can be multiply realized using physical supervenience, because it will require reproducing the *same physical relationship* everytime. Even if such reproduction is possible, it still needs substances that have a set of properties (see below).

Objection

What if the reflection on the mirror is treated as supervening on reflected light rays?

This assumes that the mirror and the object are not relevant to the image. All changes in the image are changes in reflection patterns. However, this does not help in saving the supervenience-I relation. Different objects can create the same reflection patterns, but the images would be different. So the image would change, but the physical base (the reflection pattern) wouldn't. So supervenience-I doesn't hold.

To save supervenience-I, we can assume that all reflection patterns are unique. Then a different object will create a different reflection pattern, which will result in a different image. However this solution suggests that we can't have multiple realizability, because all patterns are unique, and therefore, are not reproducible. So it looks like if we subscribe to physical supervenience, we can't have both supervenience-I and multiple realizability.

Mental material

There is another, more important, reason for this conflict: even though physical supervenience as described in the examples are multiply realizable to some extent, they can be realized *only by specific classes of substances.* For instance, magnetism can supervene only on *magnetic* materials and reflections only on *reflecting* materials. These classes of substances have *properties* that allow the supervenience relationship to exist.

By analogy, if you believe in both multiple realizability and physical supervenience in the case of the mind, then for AI to obtain, you have to postulate a class of materials on which the mind can supervene. In other words, you need some sort of mental material to recreate the mind. The only mental material we know of right now is the brain.

So postulating multiple realizability using physical supervenience is postulating a new *physical* property shared by some (or all) materials. Would the physicists agree to this?

Objection

Information processing material can be viewed as mental material.

By this hypothesis, we have to agree on information processing as a *physical* property that can be instantiated by a certain category of materials, just like magnetism. So, for instance, silicon could instantiate the mind, just as iron does magnetism.

The problem is that the category of materials that instantiate a physical property is *identified by the capacity to possess that property*. Magnetic materials are such because they can possess magnetism, reflective material are such that they can reflect. Therefore, by this hypothesis, information processing materials should first be *identified as having a mind*.

However, since the multiple realizability hypothesis does not identify such materials by *pointing them out as possessing minds*, the notion of information processing as a physical property does not help the thesis.

Supervenience and Indian Philosophy

Let me recap the points I've made:

P1: The supervenience relation that naturally supports multiple realizability is one that uses conventions. However, since this supervenience needs an observer, or a set of observers, it is not useful in solving the mind-body problem.

P2: There's another supervenience relationship, what I call physical supervenience, which is a naturally occurring relation, and doesn't need conventions. But physical supervenience poses two problems: one, it is not a strictly local relationship. For this relation to apply to the mind-body problem, the environment has to be brought into the picture. In other words, the environment has to be part of the base on which mental content supervenes. Two, it needs us to postulate some sort of mental material on which the mind can supervene.

P3: Because of the second reason above, physical supervenience is in conflict with multiple realizability, and hence AI. If A supervenes on B *by virtue of B's physical properties*, then obviously A is tied to B's physical properties, and therefore A is not multiply realizable.

Now we can ask the question whether we can save both the supervenience relationship and multiple realizability. To begin, let us ask: what kind of model can save both?

1) Obviously, given the examples, and my sympathies, a model that includes the environment.

2) But because of P3, we need a model where mental content supervenes on physical structure not by virtue of physical properties, but by virtue of mental content itself. The model should not postulate mental material, and should allow for the agent's internal physical mechanism to be anything (or at least silicon), not just the brain.

3) This implies moving away from a strictly physical supervenience relation to a more convention-like supervenience relation, where content can move freely between bases. However, we cannot use the conventional supervenience as in the dollar-coin case, because we need a model that gets rid of the need for an observer.

To suggest such a model, I take inspiration here from the Buddhist notion of *Swasamvedana* or autoreflexive awareness (as laid out in Matilal, 1986), the notion of *awareness events* that are self-aware. According to the model laid out by Matilal, an awareness event is like a lamp/light-bulb, it has the property of illuminating itself; so an awareness event in an agent doesn't necessarily need an observer/observation module to become an awareness. Though it is not specifically laid out in the model, I assume here that an awareness event is akin to a representation, and, therefore, has a link to the environment (broadly construed, i.e. including bodily states etc.). I'm encouraged to treat it this way because of the agent-environment interaction inherent in the constructivist metaphysics advocated by Buddhism, even though I am not particularly keen on traveling all the way up the metaphysical path with the Buddhist.

Now here's why I think this model could solve the problem of supervenience:

- An awareness event A supervenes on the brain+environment physical state B, so the non-reductive physicalist model of supervenience holds — any change in the mental state A is a change in the physical state B.
- The idea of these events being autoreflexive, or being "self-illuminating", gives them the ability to be multiply realizable,

realizable on any base, because *the property of being a mental event is part of the event itself, and not the physical system.*

In some ways, the model is similar to conventional supervenience; the difference is that while conventional supervenience needs an external observer and a convention to set the mental content, the autoreflexivity model, as I see it, doesn't, because the awareness and content is internal to the *event.*

In effect, the awareness event model marries physical supervenience with conventional supervenience in an elegant way. It extends physical supervenience by bringing in the environment into the supervenience relationship, but gets rid of the need for an external observer from conventional supervenience, by postulating awareness as "self-illuminating".

The awareness event model works better because it "empowers" mental content, so to speak. It considers mental content to be a bit more "powerful", having more intrinsic properties, than the passive representations used in classical AI models. Or the epiphenomenal representations postulated by eliminativist models, where the power is vested with the physical system, and the mental is a stub, an epiphenomena generated by the physical system. The awareness event model essentially moves some of the "power" to the abstract entity "event", which brings together the environment, body and the mind.

Admittedly, this is a move towards a kind of idealism, where an abstract entity "awareness event" is postulated with some internal properties. However, I consider this a plausible move — if numbers as abstract entities can have properties, awareness events can have properties as well. More closer home, an awareness event is not a lot more idealistic than the abstract inner representations postulated by most computational models. Also, I think some sort of an abstract entity is a requirement in this case, because if the property of being a mental state is a physical property, we wouldn't have needed the supervenience relation in the first place.

More worrying is the nature of the "self-illumination". How can an awareness be "self-illuminating"? It looks as if the introduction of autoreflexive awareness only substitutes a problem with a mystery. However, this need not be so. An awareness event can be viewed as something similar to action-oriented representations, as suggested by Andy Clark (1997).

Clark argues against idealised, centrally stored representations of classical AI and suggests that representations are closer to the Gibsonian notion of affordance. An affordance is "an opportunity for use or interaction which some object or state of affairs presents to a certain kind of agent." Thus, for a human a chair affords sitting, but to a woodpecker or termite it may afford something different. Thus, representations are not action-neutral, as envisioned by classical AI, but they are action-oriented.

This model of representation is very similar to the idea of a self-aware awareness event, as sketched above. The "self-illumination" property of an awareness-event is not a mystery in such a model, *it is part of the event's action-orientedness.* An action-directed awareness event is aware of itself, within the context of the action. So, the event of my hand encountering the coffee cup is aware of itself, which is why I move to the next event, raising my hand, and start drinking from the cup. Notice that this is much weaker than the claim that an agent has full-blown awareness of all her awareness–events.

The following example will illustrate the notion better. Consider the problem of cricket phonotaxis (the ability to move towards sound sources), which is being tried out in a robotic cricket by Barbara Webb and her colleagues at the University of Stirling. In real crickets, this is the process whereby a female cricket identifies a male of the same species by his song, turns in his direction, and reliably locomotes to the source. This process looks as if the cricket needs to manipulate representations to do it, i.e. the cricket reaches the male by identifying the song as coming from a direction, computing the location, and then moving towards that location.

However, the cricket achieves phonotaxis using a much more elegant mechanism. The cricket's ears are on its forelegs. Sounds are also received through another routing mechanism in the cricket's body. The combination of the two allows the cricket to fix the sound as louder on the "ear-leg" nearest to the source. There is one neuron on the leg dedicated to detect this loudness. Based on this information, the cricket automatically turns in the direction of the source. The cricket constantly changes its path based on perceived loudness of the song. This hearing-plus-locomotion system detects only the songs of the cricket's own species. This purely mechanical process can be replicated, in principle, by a robot.

Now, suppose we take a slice, grab an action frame, from this

mechanical process of the female moving towards the singing male. Each slice is oriented towards the action of reaching the mate.

- *The slice is the awareness event.*
- The orientedness towards the action of moving towards the male, the action-orientedness, provides the self-illumination.

In other words, a mental state is a slice of action. Since actions are directed, the mental states are, too. Bickhard (1998), argues something similar. According to him, situated robots have system-detectable truth value. In *inter*active robots, a representation has a truth value for the system itself. This is very similar to the notion of self-illumination.

Now, what does this mean for supervenience and multiple realizability?

- The awareness-event (slice) supervenes on the agent + environment interaction (movie). Since the slices cannot change by themselves, any change in the slice is a change in the movie.
- The awareness event is action-oriented, so it gets its directedness — it becomes an intentional state — without needing an observer.
- Since the phonotaxis process can in principle be recreated using robots (ongoing work), this supervenience is multiply realizable.

But there are outstanding questions. For instance:

1. What about mental content without action, thoughts such as "Giraffes cannot wear turtleneck sweaters"?
2. Can a slice of an agent-environment interaction (say, of a robot) become mental content just by itself? Doesn't consciousness have a role in making something mental content?
3. How come actions are directed?

The first question could have evolutionary answer, but the other two are open research questions still.

Future work

This model is not intended as a direct translation of an Indian philosophy model into Cognitive Science. It is a model that just draws inspiration from an Indian philosophy model. From an Indian philosophy standpoint, I'm treating a mental event here as an

instance of consciousness. The relationship between consciousness and mental events is quite nebulous in Indian philosophy and fitting those debates into Cognitive Science is almost an impossible task. This model is just intended as a bridge – to connect Indian philosophy and Cognitive Science. Also to connect the largely functionalist idea of supervenience with the situated cognition (and Indian philosophy) notion of mental content being episodic — i.e. mental content as being tied to the environment, instead of being passive, stored data, as suggested by classical AI models. Since the episodic, environment-directed, character of mental content is a recurring theme in Indian philosophy, I think it could be explored further to build bridges with situated cognition models in current Cognitive Science.

NOTES

[1]These are basic computational models mostly well-known to the readers of cognitive science. See for reference Stillings (1995) for a comprehensive coverage of them. Individual details can be obtained from Andersen (1990), Andersen (1983), Chomsky (1972), Fodor (1975), Minsky (1986), Newell (1990), and Pylyshyn (1984).

[2]There is of course the argument made by externalists like Millikan (1994) that the function can be "selected for" by evolution. However, in my view, "being selected for" is also a norm. The crucial difference is that the observer for whom it is a norm, namely nature, is an entity that is not considered as a subject, but is considered as being "neutral". So it is still a convention, but one apparently sanitized. I return to this point below.

[3] There is always the question of how the faceless entities come to be.

[4] Let me briefly consider Millikan's point here. Even if we grant Millikan that a representation is "selected for", she has to agree that only a certain class of materials, namely organic matter, has gone through the selection process. So I think she, too, cannot make the move to multiple realizability.

[5] Theories about the atomic structure of magnetic substances come *after* they are identified as magnetic, not before. So information-processing materials should be identified as mental materials first, *before* we launch into discussions about them having an atomic structure, if any, that makes them information processors.

[6] *Swasamvedana* is a thesis held by various Indian schools, but I find the Buddhist model interesting because it attributes autoreflexivity of awareness to discrete *events*, not to a central controller. Tom Yarnall has pointed out to me that not all Buddhists accept the *Swasamvedana* thesis. Also, questions have been raised about the suitability of using Matilal as a reference here.

[7] For my purposes, I am treating self-illumination as a property intrinsic to the abstract entity event. There could be other ways to construe self-illumination, including purely physicalist approaches. Many thanks to Dr. Ram-Prasad Chakaravarthi for pointing this out to me.

REFERENCES

Anderson, J. R. (1983). *The Architecture of Cognition*, Harvard University Press, Cambridge, MA.

Anderson, J. R. (1990). *The Adaptive Character of Thought*. Hillsdale, NJ: Erlbaum.

Bickhard, M. (1998). Robots and Representations. In Pfeifer, O., Blumberg, B., Meyer J.A., and Wilson, S.W., *From animals to animats*, proceedings of the fifth international conference on simulation of adaptive behaviour, Cambridge, Mass.: MIT Press.

Block, N. (1980). Introduction: what is functionalism? In N. Block (Ed.), *Readings in philosophy of psychology*. Cambridge, MA: Harvard University Press.

Chomsky, N. (1972). *Language and Mind*. New York: Harcourt, Brace and Jovanovich.

Clark A. (1997). *Being There: putting brain, body, and world together again*. Cambridge, Mass.: MIT Press.

Dictionary of Philosophy of Mind, available at http://artsci.wustl.edu/~philos/MindDict/

Fodor, J. A. (1975). *The Language of Thought*, Cambridge: Harvard University Press, Massachusetts.

Kim, J. (1993). *Supervenience and the mind: selected philosophical essays. Cambridge Studies in Philosophy*, Cambridge: Cambridge University Press.

Matilal, B.M. (1986). *Perception: An essay on classical Indian theories of knowledge*. Oxford: Clarendon Press.

Millikan, R. G. (1994). Biosemantics. In Stich, S. and Warfield, T. A. (eds.), *Mental representation, a reader*. Oxford: Blackwell.

Minsky, M. (1986). *The Society of Mind*, New York: Simon and Schuster.

Newell, A. (1990). *Unified theories of cognition*, Cambridge, Mass.: Harvard University Press.

Pylyshyn, Z.W. (1984). *Computation and Cognition: Towards a Foundationfor Cognitive Science*. Cambridge, Mass.: MIT Press.

Stillings, N., et al., (1995). *Cognitive Science*. Second edition.Cambridge, Mass.: MIT Press.

Webb, B. (1996). A robot cricket. *Scientific American*, 275(6), 94-99.

No Need to Have the Same Consciousness

B.A.C. SAUNDERS AND J. VAN BRAKEL

Although the title of this paper may seem ambiguous, it's not. It's meant to convey that consciousness is not a reified thing, nor a natural kind or universal. Consciousness is not something with well-determined properties – an 'essence' that might be discovered. Having abandoned the quest for an essence we might better think of a family of consciousnesses and their congeners, amongst which, and depending on what is at issue, differences and similarities may crop up and disappear with nothing at all 'given'. This is not a common view and here we shall give the barest glimpse of what we're talking about. Don't expect exciting news on neurotransmitters or the dissipative quantum model of the brain, nor some old fashioned philosophy of the sort to be found in *The Nature of Consciousness* (1997) which contains fifty 'central' papers. The editors start their Preface with: 'These are exciting times for thinking about consciousness, and this book contains a collection of essays that represents the cornerstones of contemporary philosophical thinking on the subject.' The issues raised by these articles are not going to be addressed here, but rather some presuppositions underlying them all.

Let's start with the English word 'consciousness'. According to the *Oxford English Dictionary* the first use of 'conscious' to mean 'inwardly sensible or aware', was used in a sermon given by Archbishop Ussher in 1620, who said '[I was] so conscious vnto my selfe of my great weakenesse'; but even then Hobbes, as late as 1651, continued to use the term with its earlier sense of 'shared knowledge'. Similar French and German novelties appeared at roughly the same time. To begin with then we would like the close relation of consciousness to knowledge, especially in the sense of *shared* knowledge, to be kept in mind.

The 'modern' use of the term 'consciousness' can be traced to the mongrel root metaphor of Cartesian dualism and the think/

feel dichotomy. Due only to this starting point does the issue of *the location* of consciousness arise – focussed as it is in the brain. That this postulate is a modern invention of the western tradition is easily shown by considering its pre-modern precursors. In classical Greek writings there is no term that even roughly translates as either 'mind' or 'consciousness'. Instead we find the term *psyche* – often, though misleadingly translated as 'soul'. Though *psyche* changed considerably in sense between Homer and Aristotle, it kept one highly significant connotation throughout: to be alive is to have *psyche*, to have *psyche* is to be alive. Hence bamboo and spiders have *psyche*. They have it just as much as humans do, even though they have it in a less sophisticated form. What it is to be human requires reference to the capacities for language and reason, which also belong to the powers of *psyche*. Moreover there is constant feedback between all levels of the physical realisation of *psyche*. A similar view can be found in the writings of William James or Charles Peirce and in various non-western traditions. An obvious merit of *psyche* is that it directs attention to the powers of organisms. Emphasis on the conscious mind in the Cartesian tradition, tends by contrast to highlight events, sensations, or 'ideas' – countable things, rather than powers; i.e. it focuses on the statics of properties, rather than on the dynamics of relations and actions.

The schism Cartesianism forced between 'conscious' and 'non-conscious' might be called the 'Cartesian catastrophe'. 'The mind', in other words, was hived-off from the body. Only in the second half of the twentieth century has the body however re-appeared, via writings of philosophers such as Merleau-Ponty, and as a consequence of gender interests. It became second nature to assess the role of each side of the dichotomy and to cast theories in terms dictated by it; hence the main worry has been concerned with how to bring the two realms together. It is therefore via the notion of *psyche* that the relation of consciousness to life, powers, and reason, as well as to shared knowledge needs to be kept in mind.

Let's now look briefly at Chinese philosophy. Xin (1984:78) writes: 'there was, strictly speaking, no concept or special term in ancient Chinese philosophy which could now be translated as "consciousness" . . . but that does not mean that the ancient Chinese philosophers did not discuss the "problem of consciousness"'. None of this however implies they were discussing 'consciousness' as currently understood in the dominant literature. Wilkes (1988), quotes an unpublished report of the Chinese Academy of Science saying that *yi shi* comes closest to a translation of consciousness. According to dictionaries the two components, *yi* and *shi*, originally had much the same meaning: knowing, or remembering. In old Chinese *yi* seems to have been closer to tacit, implicit knowledge than to conscious, front-of-mind knowledge. Today *yi* includes (a) meaning or implication; (b) wish, desire, or intention; (c) conception, idea; (d) anticipation or expectation. *Shi* is primarily knowing, or knowledge. Hence both *yi* and *shi* are highly cognitive. More recent meanings of *yi shi* seem to cover the various mental processes or products of "the mental" such as sensation, perception, representation, concepts, judgement, inference, abstraction, generalisation, as well as feeling, emotion, will and personality – i.e. it is aligned to what may be found in any cognitive psychology text book.

In *yi shi* we see again the close connection between consciousness and knowledge. Of course this is not surprising, but in English the two terms would not normally be considered near-synonyms. Compare this with what Kern (1992: 264) says about Xuanzang, a Chinese Buddhist monk and thinker (of the seventh century A.D.): '[he] continues to conceive the object as a *condition* for the arising of the knowledge (or consciousness) of it.' No principled distinction is made between knowledge of, and consciousness of, an object therefore.

We now turn briefly to the vast literature on consciousness in Indian philosophy. Issues that are here addressed include:

- is consciousness consciousness of some object or not (i.e. does it have intentionality)?
- is consciousness constitutive or foundational (i.e. is it the ground of everything)?
- how are "pure", "phenomenal" and "transcendental" consciousnesses to be balanced or harmonized?
- what is at issue when one talks of "levels" of consciousness?
- what is the relevance of mystic experiences (if any)?
- should mind be viewed as a material faculty (and if so, in what sense)?
- does the non-duality advocated by some Indian schools of thought avoid the Cartesian mind-body problem?

What is noteworthy is that all writers consulted assume consciousness to be one thing. Or if different types or levels of consciousness are introduced, they are held to be a sub-classification of one

"essential" domain. The contested issues are disagreements about what the best theory or essential properties of consciousness are. It is rare to find the fluidity or plasticity of concepts, their normative proprieties, or their use brought to the fore.

Moreover, many writers tend to assume that western philosophy should be the measure of all things. Mohanty, for example, rejects certain schools in Indian philosophy simply because they are not consistent with his Husserlian phenomenological starting assumptions. In *The Nyaya Dualist Tradition* (1999) Chakrabarti translates the Sanskrit term *guna* as "quale", though more usually it is "quality" or "property". Whatever *guna* might mean however, it would be implausible if it could substitute for "quality" or "property". It would be even more implausible if the latter were understood as modern technical philosophical terms. Narain (1961) already argued against the established practice of translating *guna* as 'quality' and proposed "attribute" instead in order to cover "quality", "quantity" and 'relation':

- difference, conjunction, disjunction, remoteness, and proximity are *gunas* – in western terminology these would be called "relations" not "qualities";
- in the western tradition motion is usually considered a (primary) *quality*, but in the "*guna* tradition" it is assigned a place of its own (beside the *gunas*)
- *guna* is usually taken to be unrepeatable which is not typically associated with quality (it's on this basis that Chakrabarti favours qualia as a translation of *gunas*).

However to suggest that *guna* means "quale" is strange, because "quale" doesn't mean anything in ordinary English. It is a purely technical term invented by philosophers of a particular persuasion – those of a different persuasion denying it sense. To assert that *guna* means quale is an example of language going on holiday, as Wittgenstein would say. Surely the writers in the Nyaya tradition who reflected on *guna* weren't taking a position in a philosophical debate that emerged only in twentieth century analytic philosophy? Why not just use the term *guna* in say, English texts, without translating it, and try to elucidate its polysemies with as many examples and analogies as possible?

A view shared by many philosophers, one which is a fundamental pillar of the scientific approach, and which has insinuated itself into

every nook and cranny of the planet, is the distinction between 'inner' and 'outer'. This is a kind of root metaphor for such dichotomies as subjective-objective, fact-value, literal-metaphorical, and so on. However if one looks around (i.e. across a variety of human life forms in different times and places), it becomes less clear what belongs to what (cf. Table 1). Whatever is considered "inside" varies so greatly it cannot be theorised independently of what is "outside"; so-called boundaries between inside and out are blurred. Neither "inner" nor "outer" nor their separation warrants essentialisation, or for that matter any attention at all. For example, for the Nuer (of the Sudan) "belief" cannot be established as a distinct inner state (Needham 1972). Needham could not find anything corresponding to the popular modern conception of mind among the Nuer, despite what he considered a very elaborate indigenous "psychological" vocabulary.

Similar remarks apply to notions such as "emotions", "feelings", "experience", "self". They can't be pinned down; rather, they go all over the place. It's only for lack of a better discursive horizon that we gesture with such limited western terms as "emotion" or "feeling" in the direction of aspects of human life forms. For example, it's not "obvious" that Sanskrit *bhava* means "emotion" (passing over the fact that "emotion" can't be regarded as a single "thing' in English either).

"Experience" has multiple meanings in scholarly English (at least four if not six according to Feenberg 1999), each of which would fragment into a variety of usages on further analysis. Havoc has been caused by assuming "experience" to be one thing in western philosophy. Varieties of "self" too are well known. In many societies the boundaries between self, other, and cosmos are less distinct and/or the multiplicity and separability of components are recognised. In the western tradition too there are those who emphasise that persons (or selves or positioned subjects) should primarily be seen as intersections of social relations, or points of intersection between the subjective and the social. There are views of human personality that assume a fluid complex of functions or living forces characterised by intrinsic intelligence and co-ordinated by a dynamic structural principle. To give one concrete example, the Advaita school in Indian philosophy uses a threefold classification of the self, depending on three construals of consciousness: *jiva* (individuated self with its own parameters of reflexive occurrence), *atman* (auto-

reflexivity – sometimes glossed as "self"), *brahman* ("general consciousness": a reflexive, singular, irreducible, universal entity). Advaita recognises that there is a tension between first and last. However that tension is resolved however, their three-fold self is not that of the western individual. The Advaitin's "self" is "the sheer fact of auto-reflexivity, whose identity lies, paradoxically, in the non-individuated universality of a singular yet general entity, *brahman*." (Ram-Prasad 2001: 389).

As to emotions and their congeners, they have been located within the mind, within the body, occurring in or to others, as things coming from outside, located in events, associated with particular contexts, and so on. For example, Samoans will say that they are aware of no particular bodily feeling that accompanies emotions, yet Samoan people laugh, cry, flush, shout, and so on, in situations that are interculturally recognisable as affectively moving (Gerber 1985). The origins of such movements have been located in the heart, lungs, liver, stomach, intestines, the eye, or in 'pseudo-organs,' such as souls which are detachable from the body (cf. Table 1).

According to some sources, in Hindi and Dravidian languages, no particular word translates as 'to feel' (e.g. Lynch 1990: 11): anger comes *to* one, one loves or envies *with* another; others (e.g. Wierzbicka 1992) argue that the Hindi word for 'feel' is *manas* – a word that is also translated as awareness, mind (as the governor of the senses) or self (more social than awareness). Would it help to ask Hindi and English speakers to resolve the issue for us? No. If we were to conduct a questionnaire with bilingual people, we would predict no consensus on whether *manas* could be used substitutionally for all uses of 'to feel' or 'feeling'.

Consider too the complex notion of *rasa*. Amongst many descriptions are the following:

- the consistent and harmonious expression of the human spirit in feelings, sentiments, and emotions through various modes of art creation (as distinct from *dhvanti*, glossed as 'sentiment, sense')
- the core emotive content of an art work;
- the relation between word and meaning in poetry;
- a group of 'aesthetically distanced' emotions;
- interaction between a disposition and an object culminating in a particular expressed feeling;
- something derived from direct sensual perception or experience, but more pristine and rarefied than any feeling;

TABLE 1. Varieties of "emotion talk" and the problem of locating "the cognitive" and "the affective".

Balinese	*keneh*	the 'feeling-mind' ('who can think but with their feelings'); emotion is bound in a conception of social obligation and balance in the cosmos.
Chewong	*rus*	'liver', the seat of thoughts and feelings; using expressions such as 'my liver is good, tiny, forgot' for feeling fine, ashamed and forgetting something; distinct from the *ruway*, the essence of being, situated in the chest.
Fiji Hindian	*bhaw*	emotion, gesture, display (with aesthetic overtones); *bhaw* 'feelings' are not experiences or viewed as internal states, but located in events.
Giriama	*ini, moyo,*	these refer to the physical liver, heart, and eye; both *ini* and *dzitso moyo* are the 'seat' of innermost sentiments, feelings, desires, etc.; greed and envy also originate from the eye; selfhood is located in the heart; the core of an argument in the liver.
Ifaluk	*niferash*	'our insides', subdivided in *nunuwan* and *tip* which distinguish between socially standard and idiosyncratic 'processes;' fuse thought and feeling and don't distinguish desire and intention.
Japanese	*jodo*	usually translated as 'emotion', but includes considerate, motivated, lucky, calculating.
Malay	*hati*	'liver', which is the source of the passions (anger, envy, cheerfulness, love, etc.); *sakit hati* 'liver sickness' =? mental disorder.
Maori	*ngakau*	the 'mind' as an undifferentiated whole of cognition, emotion and volition, located in the intestines; other organs (*manawa* 'connected with breathing', *mauri* 'life principle', *wairua* 'spirit of person that can move about', and others) may also experience emotions.
Mangap-Mbula	*kete*	'liver'; term used for talking about all kinds of feelings or states linked with feelings.
Minang-kabau	*raso ati*	usually translated as 'emotion' but includes 'seriousness', 'honesty', 'indecision'.
Newar and	*nuga*	'the sacred mind', located in the heart, the seat of cognition, memory, perception, emotional experience inhabited and animated by 'a moral god'.
Samoan	*loto*	in general: 'depth'; as the seat of emotions physically located in a person's body, but its precise site is ambiguous.
Tahitian	*'a'au*	'intestines'; where emotions are located.
Temiar	*hup*	the heart, where feelings, thoughts, awareness, and memory are placed, but not vocalised expressions which are located in the 'head soul' (*rewaay* 'animating principle').

- an autonomous meta-emotion, a *sui generis* form of consciousness;
- a refined mind-body experience; in particular *srngara rasa* - divine erotic love/play not directed at humans (sexual desire abstracted from all sexual implication); also glossed as "a culturally constituted emotion that is embodied thought" or "the passion of life";
- that which goes beyond the crude world of materiality to discover *rasa's* aesthetic purity as a manifestation of supreme love;
- that which points to the basic harmony of the universe behind its apparent multiplicity in ordinary life;
- that which is expected to be conducive to the basic goals of life;
- "juice, sap, liquid, to taste, to savour, to sample" – "*rasa* emotion talk" is inseparable from metaphors and metonyms of food practices;
- an analogical intermingling of the notions of "the right gestures", "the right emotions", "the right flavours of a dish", "the right divinised dance" and such like.
- Clearly a lot needs to be said before one can get an inkling of what *rasa* is or is like, but it won't help much to insist on an inner/outer distinction.

In a longer paper first person authority would have to be addressed; that is, the issue of whether or not we can be certain of the contents of our own mind in a way we cannot be of other kinds of knowledge. Instead we limit ourselves to one comment. Consider the following types of knowledge: of a shared "external" world, of other minds, of one's own mind. Each is objective in its own way; none is less or more privileged. Why? Because they can't do without one another: "objectivity" is a "co-production" of all three. Other minds can only be "accessed" via "their" interaction in a shared world and can only be interpreted via "my" meanings. But I can only have my meanings if I am with others with whom I share a world. Further, knowledge of a shared world and knowledge of other minds are necessary to "create" mind in "new" persons.

While mind, consciousness, and their congeners are in some sense important, they cannot be definitively theorised, because they are not the kind of *object* necessary to a scientific theory. Moreover consciousness is syncategorimatically tied to reason and rationality, to shared knowledge, to meaning, mind, self, soul, *psyche*, *yi shi*, *gunas*,

manas, *rasa*, and so on – none of which lend themselves to precise, let alone complete description. If one tries to reify reason or consciousness, it will turn out that that which one wants to capture in terms of definitions, theories, and essences, transcends what is being theorised.

Neither matter nor consciousness need be taken as inherently scientific and quantitative. This is already found in Aristotle and takes away the necessity of concentrating on the brain as the organ of thought. Rather one might regard the whole body in its fleshly, organic nature as thinking, relational substance. Thus our body uniquely binds us into the world because the world is the way it is because our bodies are the way they are. On this basis it is not consciousness that is basic, but engaged agency. It's not that mind and matter don't exist, but that to talk about them in that way is a derivative, parochial approach because dualism and mechanism are ontologies of disengagement, and so give a distorted understanding of existence and the world. It is being participants in a shared practical world that gives us a window on ourselves and reality, including transcendental reality (were there to be one).

Some might claim mystic access to consciousness: but each mysticism would find its own, dependent on the form(s) of life with which it is engaged. Neither science, nor mysticism has the capacity to help us find *the* essence of consciousness. A more modest view would be to aim for a perspicuous presentation of the variety of ways of thinking and talking about the "family" of consciousnesses, inner and outer, emotions, rationality, and so on. These ways are not those of isolated solipsistic individuals, but rather embedded in form(s) of life (of which scientific forms occupy but a few neighbourhoods).

To establish what congeners of consciousness have in common would require extensive expositions of the form of life in which they have their place. Different ideas about consciousness slide into one another and vary with time and place, while their variations depend too upon one's situatedness. While there may be many "similarities" (depending on what is at issue), there is no one true taxonomy or essence. If the distinction between thought and feeling, or body and mind, or the moral and the personal is made differently, or not at all, or if it is assumed that cognitive and volitional aspects of mental life are intimately related to the affective or emotional, there will be a profound effect on how consciousness is approached.

A proper scientific concept, one might suggest, could be bootstrapped out of common sense notions. Naturalised philosophers of science for example might say that instruments such as microscopes and PET-scans provide pictures of reality, because they extrapolate from the capacity of the human eye. But this argument is specious: the needed analogy between eye and microscope isn't there. An eye isn't an instrument; it has a *psyche*. But were the argument to be accepted along with the claim to reality of salient properties of neurological events, it would still not be clear what was being bootstrapped into what, and what abstracted away.

In being critical of consciousness as a well-determined "thing", we're not saying that there's no self-consciousness. But nor are we saying that there's no *psyche, yi shi, manas*, or *gunas* either.

Provided it is not assumed that introspection or self-awareness is synonymous with consciousness, there is no harm in talking about "it". Consciousness is one result or effect of *Mitsein*, of being inducted into a human form of life. But such awareness only *derivatively* and *locally* consists of an awareness of the uniqueness of "the self". It may equally consist of a consciousness of panhuman solidarity and sameness (or of something similar, for example restricted to a local "we"). It may look inward or outward: *inwardly* we may find souls, psyches, minds, undivided selves or a series of faculties, or something else again; *outwardly* we may find awareness of social insight, self-knowledge, mystical communion, technical discoveries, elimination of individualistic thought and striving for moral determinations (Lutz 1985). And so on. The limits of awarenesses of consciousnesses are the limits of the manifest form(s) of life or lifeworld(s).

These manifest form(s) of life should be understood at the same time in the singular and plural, as local and universal, as empirical and transcendental. Both one and many human forms of life may emerge and fade away – similarities, (re)semblances, as well as variations, contrasts and distinctions, pop up and disappear (depending on what is at issue in the local *Mitsein*, without there being a common core or incommensurability).

That different form(s) of life may in some sense appear 'similar' is no reason to suppose that there is a core, or some essence, rationality, prototype, DNA-structure, transcendental subject, transcendental *We*, or *Dasein*, which of necessity, all share. The argument is analogous for consciousness. There's no *practical* reason to worry about living in totally incommensurable worlds; there's no

reason to worry about the problem of other minds – if there hadn't been other minds there wouldn't have been a mind in the first place. However, from these well supported empirical (or if you like, transcendental) facts it doesn't follow that there *has to be* a shared core or essence of human behaviours; or a shared *lingua mentis*, mind, or "consciousness", or any other preconceptual, cognitive, affective, kinaesthetic, or communicative universal structure that guarantees the success of communication, specifying pancultural basic emotions, colours, directions, universal principles of empathetic understanding, ideal speech situations or whatever.

This doesn't mean that the word "universal" becomes meaningless. It merely means that "universal" is not a *universal*. How the notion of universality (or of rationality or morality) is understood is always contestable again and again, and there are no transcultural criteria for making a final judgement about which contestations are "appropriate" and which are not. Put differently: if there is any ground for using the term "universal" *with justice*, then this should leave open the possibility that this local concept of universality is contested by those who have not participated in its formation (Butler 1996).

In stressing the relevance of the manifest life form(s) or life-world(s), we're not taking an anti-science position. On the contrary we're merely saying that *if* (and only if) the question of the tension between the manifest and scientific image is raised, it's *not* the scientific image that can claim priority over the manifest life form(s) – the latter expression to be understood interculturally. It is *not* the scientific image that can have the last word on "consciousness". What instead we should be looking at is a perspicuous presentation of the variety of consciousnesses throughout the time-space of humans. The place to start is not then with a solipsistic investigation of one's own thinking and experiences, be they scientific or mystic. It is rather to focus on the variety of manifest form(s) of life or lifeworld(s).

REFERENCES

Aklujkar, A. (2001) The word *is* the world: nondualism in Indian philosophy of language, *Philosophy East & West*, 51: 452-473.

Balslev, A.N. (1993) The influence of phenomenology on J.N. Mohanty's understanding of 'consciousness' in Indian philosophy, pp. 245-254 in *Phenomenology: East and West: Essays in Honor of J.N. Mohanty* (F.M. Kirkland and

D.P. Chattopadhyaya, eds.), Dordrecht: Kluwer.
Bhattacharyya, S. (1992) Phenomenology and Indian philosophy, pp. 44-76 in *Phenomenology and Indian Philosophy* (D.P. Chattopadhyaya, L. Embree and J. Mohanty, eds.), New Delhi: Indian Council for Philosophical Research.
Block, N., O. Flanagan and G. Güzeldere (1997) *The Nature of Consciousness: Philosophical Debates*, Cambridge Mass: MIT Press.
Bogoras, W. (1925) Ideas of space and time in the conception of primitive religion, *American Anthropologist*, 27: 205-266.
Busby, C. (1997) Permeable and partible persons: a comparative analysis of gender and body in Southern India and Melanesia, *Journal of the Royal Anthropological Institute*, 3: 261-278.
Butler, J. (1996) Universality in culture, pp. 45-52 in *For Love of Country: Debating the Limits of Patriotism* (J. Cohen, ed.), Boston: Beacon.
Chakrabarti, K.K. (1999) *Classical Indian Philosophy of Mind: The Nyaya Dualist Tradition*, Albany: State University of New York Press.
Chethimattam, J.B. (1970) Rasa, the soul of Indian art, *International Philosophical Quarterly*, 10: 44-62.
Davidson, D. (1987) Knowing one's own mind, *Proceedings and Addresses of The American Philosophical Association*, 60: 441-458.
Feenberg, A. (1999) Experience and culture: Nishida's path 'to the things themselves', *Philosophy East & West*, 49: 28-44.
Gerber, E.R. (1985) Rage and obligation; Samoan emotion in conflict, pp. 121-167 in *Person, Self, and Experience: Exploring Pacific Ethnopsychologies* (G.M. White and J. Kirkpatrick, eds.), Berkeley: University of Calfornia Press.
Kern, I. (1992) Object, objective phenomenon and objectivating act according to the 'Vijñāptimatratasiddhi' of Xuanzang (600-664), pp. 262-269 in *Phenomenology and Indian Philosophy* (D.P. Chattopadhyaya, L. Embree and J. Mohanty, eds.), New Delhi: Indian Council for Philosophical Research.
Lynch, O.M. (ed.) (1990) *Divine Passions: The Social Construction of Emotion in India*, Berkeley: University of California Press.
Lutz, C. (1985) Ethnopsychology compared to what? Explaining behavior and consciousness among the Ifaluk, pp. 35-77 in *Person, Self, and Experience: Exploring Pacific Ethnopsychologies* (G.M. White and J. Kirkpatrick, eds), Berkeley: University of Calfornia Press.
Lutz, C. (1988) *Unnatural Emotions: Everyday Sentiments on a Micronesian Atoll and Their Challenge to Western Theory*, Chicago: The University of Chicago Press.
Marsella, A.J., G. DeVos, and F.L.K. Hsu (eds.) (1985) *Culture and Self*, New York: Tavistock.
Mohanty, J.N. (1988) Phenomenology and Indian philosophy: the concept of rationality, *Journal of the British Society for Phenomenology*, 19: 269-280.
Narain, H. (1961) Finding an English equivalent for '*guna*', *Philosophy East & West*, 11: 45-51.
Needham, R. (1972) *Belief, Language, and Experience*, Oxford: Blackwell.
Patankar, R.B. (1980) Does the *Rasa* theory have any modern relevance?, *Philosophy East & West*, 30: 293-303.
Potter, K.H. (1954) Are the Vaisesika '*gunas*' qualities?, *Philosophy East & West*, 4: 259-264.
Ram-Prasad, C. (2001) Saving the self? Classical Hindu theories of consciousness and contemporary physicalism, *Philosophy East & West*, 51: 378-392.

Rosaldo, M.Z. (1984) Toward an anthropology of self and feeling, pp. 137-157 in *Culture Theory: Essays on Mind, Self, and Emotion* (R. Shweder and R. LeVine, eds), Cambridge: Cambridge University Press.
Saunders, B.A.C. (1993) Disenshrining the Cartesian self, *Behavioral and Brain Sciences* 16: 77-78.
——. (2002) Aristotle's non-reductive, anthropocentric materialism, pp. 189-200 in *Theories, Technologies, Instrumentalities of Colour: Anthropological and Historiographic Perspectives* (B.A.C. Saunders and J. van Brakel, eds.), Maryland: University Press of America.
Shweder, R.A. (1993) The cultural psychology of the emotions, in *The Handbook of Emotions* (M. Lewis and J. Haviland, eds), New York: Guilford.
——., and R.A. Levine (1984) *Culture Theory: Essays On Mind, Self, and Emotion*, Cambridge: Cambridge University Press.
van Brakel, J. (1992) The complete description of the frame problem, *Vivek* (Bombay), 5 (3) 11-16.
——. (1994) Emotions as the fabric of forms of life: A cross-cultural perspective, pp. 179-237 in *Social Perspectives on Emotion* (W.M. Wentworth and J. Ryan, eds), Vol. II, Greenwich USA: JAI Press.
White, G.M., and J. Kirkpatrick (1985) *Person, Self, and Experience: Exploring Pacific Ethnopsychologies*, Berkeley: University of California Press.
Wierzbicka, A. (1992) *Semantics, Culture and Cognition: Universal Human Concepts in Culture-Specific Configurations*, New York and Oxford: Oxford University Press.
Wilkes, K.V. (1988) ————, yishi, duh, um, and consciousness, pp. 16-41 in *Consciousness in Contemporary Science* (A.J. Marcel and E. Bisiach, eds), Oxford: Clarendon Press.
Xin, Ru (1984) Mind and consciousness in Chinese philosophy: a historical survey, pp. 77-86 in *Analecta Husserliana* (A.-T. Tymieniecka, ed.), Dordrecht: Reidel.

Relativism, Self-Referentialilty and Beyond Mind

KUNDAN

Challenges to objectivity

The birth of science was buttressed by a philosophy that has been called naÔve Realism, which contends that there is an objective reality independent of the observer. In other words, objectivity was the cornerstone of the Enlightenment or the Modern era where it was presumed that science following a definite methodology would be able to solve all the mysteries of the world. The unarticulated assumption was that there is a world that existed separate from the individual and that it can be understood by wresting out its secrets by a rational, unbiased and value-free observer. Consequently, the philosophy of realism created dualism such as subject and object and sharp divisions like objective reality and subjective feelings. This philosophy was the bedrock of scientific investigation until quantum physics began to complicate the matter.

In the study of high-energy particles, it has been found that particles cannot be understood as isolated entities but only in the context of their preparation and measurement. This means that the Aristotelian or the Newtonian idea of fundamental basic building blocks does not hold water anymore. Further, the classical distinction between subject and object – which was a natural outcome of the philosophy of Realism – has become vague as an observer has been found to be an integral part of the experiment. How an experimenter has set up an experiment and the measurement that he or she has decided to make determine the result of an experiment to a large extent. Thus, Capra (1992) observes:

> The human observer constitutes the final link in the chain of observational process, and the properties of any atomic object can only be understood in terms of the object's interaction with the observer. This means that the classical ideal of objective description of nature is no longer valid. (p. 78)

With the advent of the Relativity theory of Einstein, space and time, which appear to us as absolutes in our everyday experience, have been rendered relative with respect to the observer. The claim of the realists that objects like tables, chairs, bags, stones, statues, etc. have absolute existence also does not hold true in the light of the theory of relativity, for it has been shown that the length of an object – consequently its shape too – is dependent on its motion with respect to the observer. The length of a rod shortens as its motion increases with respect to the observer. Modern physics has also exploded the myth of an absolute linearity of time. Time in the theory of relativity has a meaning only with respect to a frame of reference, for as the velocity relative to the observer increases, time intervals increase. This means that the clock of the frame of reference of the observer slows down. In other words, time for two individuals moving at different velocities presents a different meaning.

The subject-object dichotomy has been critiqued from various perspectives, which has further problematized the existence of an unbiased, rational, value-free and objective subject. Language itself has been found to compound the problem of non-interactive subject and object. This happens because language does not only describe events, but also creates a cosmology, a worldview that influences the thought, behavior and perception of mankind. When a child begins to learn a language, the worldview of her ancestors is passed onto her. The pedagogic procedures used "both shape the 'appearance', or 'phenomenon', and establish a firm connection with words, so that finally the phenomena seem to speak for themselves without outside help or extraneous knowledge" (Feyerabend 1993: 57). The human mind begins to take many facts of life as givens, and the entire process may be totally unconscious. Her worldview begins to create what she may observe. Also, in order to be unprejudiced, one will have to abandon language itself, which will remove all ability to perceive and to think, as a consequence of which the practice of science will stop before it begins. Writes Edward Sapir:

> Human beings do not live in the objective world alone, nor alone in the world of social activity as ordinarily understood, but are very much at the mercy of the particular language that has become the medium of expression of that society. It is quite an illusion to imagine that one adjusts to reality essentially without the use of language and that language is merely an incidental means of solving specific problems of communication or reflection. The

fact of the matter is that the "real world" is to a large extent unconsciously built up on the language habits of the group. . . .We see and hear and otherwise experience very largely as we do because the language habits of our community predispose certain choices of interpretation. [Cited in Whorf 1962: 134]

The history of science: objectivity demystified

The publication of *The Structure of Scientific Revolutions* by Thomas Kuhn has posed the most serious challenge to the idea of objectivity that science claims to achieve. Apart from stating that science is chiefly a paradigm-based activity, he has made it explicit and clear that science, like any other human activity, is a social activity which affects and gets affected by the milieu in which it is embedded and guided by sociological, economic, historical and political forces.

Kuhn (1970) points out that scientific practice is shaped by deep assumptions of the worldview of which the scientist may be unaware. For meaningful research to take place, the scientific community must agree upon the goals, the methodologies, and the valid subject matter in the context of research. The agreement on all these issues would constitute a framework or a paradigm within which the investigation of nature can take place. The paradigm has two components – *disciplinary matrix* and *shared exemplars*. The *disciplinary matrix* consists of a certain fundamental set of assumptions that are often unstated and not subject to empirical test. These assumptions form the basis for testing specific hypotheses. For example, reductionism states that the world can be understood by breaking it into smaller units until we arrive at a set of fundamental units. This is an assumption that is not going to be subjected to any kind of an empirical test, and thus constitutes a portion of disciplinary matrix of scientists who adhere to this belief. Before scientific research begins, the community of scientists reach a consensus as to what are the essential components of the universe, what are the valid questions of research, what is "scientific" and what is not, etc. In other words, scientists know *a priori* what the universe is like. As an example, while analyzing how Descartes influenced what was admissible in the scientific canon and what was not, Kuhn (1970) writes:

[A]fter the appearance of Descartes's immensely influential scientific writings, most physical scientists assumed that the universe was composed of microscopic corpuscles and that all natural phenomena could be explained in terms of corpuscular shape, size, motion, and interaction. That nest of commitments proved to be both metaphysical and methodological. As metaphysical, it told scientists what sort of entities the universe did and did not contain: there was only shaped matter in motion. As methodological, it told them what ultimate laws and fundamental explanations must be like: laws must specify corpuscular motion and interaction, and explanation must reduce any given natural phenomenon to corpuscular action under these laws. More important still, the corpuscular conception of the universe told scientists what many of their research problems should be. (p. 41)

And again,

Paradigm functions by telling the scientist about the entities that nature does or does not contain and about the ways in which those entities behave. That information provides a map whose details are elucidated by mature scientific research. And since nature is too complex and varied to be explored at random, that map is as essential as observation and experiment to science's continuing development. Through the theories they embody, paradigms prove to be constitutive of research activity....In learning a paradigm the scientist acquires theory, methods, and standards together, usually in an inextricable mixture. (p. 109, my italics)

The other component of a paradigm is *shared exemplars* – the models for investigating new problems which include the methodology for pursuing the research. The disciplinary matrix and shared exemplars, by constituting the paradigm, unconsciously train a researcher to approach a problem in a specific way which gradually becomes his or her natural way. Through the lectures, textbooks and laboratory exercises, the members of a scientific community l earn to practice their trade. The pedagogy used orients the mind to perceive the universe as composed of certain entities and laws, the existence of which cannot be doubted under any circumstances. No one doubts the existence of electrons today just as the existence of ether was not doubted a couple of centuries ago. In this vein, Leahey (1991) writes:

> Neither source of data is comprehensible without training, yet once the scientist learns to interpret them, he or she will see them in those ways and no others. Thus training can act as a set of blinders, keeping the scientist from seeing in new ways. All observation and perception – whether scientific or not – is a matter of interpretation as numerous psychological examples have shown. (p. 14)

Weber (1946) similarly contests the idea that science can ever be free from suppositions, for it presupposes that the rules of method and logic are valid—a notion which cannot be tested by scientific means. Further, facts are meaningless and neutral in themselves; they become facts when interpreted against a theory comprising *a priori* categories. For example, the measurements made with the Atwood machine would have meant nothing in the absence of Newton's *Principia*. Varied meanings can be ascribed to the same data. What once was a Leyden jar became a condenser, as there were changes in the electrical paradigms. Elucidating how the same entity can be interpreted in different ways under the influence of different paradigms or theories, Kuhn (1970) writes:

> An investigator who hoped to learn something about what scientists took the atomic theory to be asked a distinguished physicist and an eminent chemist whether a single atom of helium was or was not a molecule. Both answered without hesitation, but their answers were not the same. For the chemist the atom of helium was a molecule because it behaved like one with respect to the kinetic theory of gases. For the physicist, on the other hand, the helium atom was not a molecule because it displayed no molecular spectrum. Presumably both men were talking about the same particle but they were viewing it through their own research training and practice. (pp. 50-1)

In short, Kuhn has shown that science is not as rational and objective as it had been supposed. Indeed, scientific rationality is a matter of consensus. It involves unexamined biases and social interests like fame, fortune, love, loyalty and power of the investigator. A choice of one paradigm over another may be induced by inner psychological causes or other sociological ones that cannot be defended by appealing to the office of reason. More often than not, scientists following the same norms of disinterestedness, objectivity and rationality arrive at different conclusions. The history

of science reveals that there are many competing theories before one paradigm becomes dominant and all of them have arisen from experimentation and observation. Comments Kuhn (1970):

> [E]arly developmental stages of most sciences have been characterized by continual competition between a number of distinct views of nature, each partially derived from, and all roughly compatible with, the dictates of scientific observation and method. What differentiated these various schools was not one or another failure of method – they were all "scientific" – but what we shall come to call their incommensurable ways of seeing the world and of practicing science in it. Observation and experience can and must drastically restrict the range of admissible scientific belief, else there would be no science. But they cannot alone determine a particular body of such belief. An apparently arbitrary element, compounded of personal and historical accident, is always a formative ingredient of beliefs espoused by a given scientific community at a given time. (p. 4)

The history of science also demonstrates that scientific knowledge is temporally relative; what was considered scientific once, has been rejected as superstition later. By the same token, what today constitutes scientific knowledge, which has been extracted from nature by subjecting it to repeated investigation, may turn out to be error tomorrow under the influence of a different paradigm. Kuhn (1970) states:

> [H]istorians confront growing difficulties in distinguishing the "scientific" component of past observation and belief from what their predecessors had readily labeled "error" and "superstition." The more carefully they study, say, Aristotelian dynamics, phlogestic chemistry, or caloric thermodynamics, the more certain they feel that those once current views of nature were, as a whole, neither less scientific nor more the product of human idiosyncrasy than those current today. If these out-of-date beliefs are to be called myths, then myths can be produced by the same sorts of methods and held for the same sorts of reasons that now lead to scientific knowledge. If, on the other hand, they are to be called science, then science has included bodies of belief quite incompatible with the ones that we hold today. (p. 2)

A committed believer in science would say that the above stated

phenomenon has taken place because science is cumulative and scientists have refined their theories in an effort to come closer to a truer and more accurate interpretation and description of nature. Kuhn disagrees and contends that instead of science being cumulative, it is revolutionary. A change in the paradigm changes the worldview of the scientist; or, in other words, the world comes to be viewed differently by the scientist. It involves a "reconstruction of the field from new fundamentals, a reconstruction that changes some of the field's most elementary theoretical generalizations as well as many of its paradigm methods and applications"(Kuhn 1970: 85).

Kuhn holds that it is difficult to demonstrate the superiority of one paradigm over another purely on 'logical' argument. The primary reason is that the proponents of the rival paradigms subscribe to a different set of standards, metaphysical assumptions and methods. The rival paradigms are so incommensurable that no appeal to 'rationality' can settle the issue as the practitioners live in different worlds. Feyerabend (1976) writes:

> Transition to criteria not involving content thus turns theory choice from a rational and "objective" and rather one dimensional routine into a complex discussion involving conflicting preference and propaganda will play a major role in it, as it does in all cases involving preferences. [Cited in Chalmers 1982: 138].

The contention of the realists that the true basis of scientific knowledge should proceed from an unbiased and unprejudiced mind, is further rendered absurd by the practice of the scientists to consider only such data which is relevant to his or her research. Since the idea of relevant and irrelevant is always present during the course of investigation, the possibility of an unbiased and unprejudiced observer takes a back seat. The investigator or scientist cannot but be an integral part of the research work and his or her subjectivity is bound to play an instrumental role in the outcome of the research. Thus, it can be safely said that the data that is generated by the scientist is not objective but collected within the larger framework of theory. It does not have an independent existence; rather it is constructed within the confines and boundaries of a theory. In other words, data is theory-laden and objectivity is the last thing that scientists should claim.

Relativity and the paradox of self-referentiality

The aforementioned arguments not only undermine the possibility of an objective truth but also demonstrate that the relationship between the subject and object is very vague, and it is difficult to differentiate them into two distinct entities.

It can also be deduced from the previous pages that all forms of scientific knowledge are relative with respect to individuals, time, and paradigms. Incidentally, this is a statement suggesting an absolute truth, which culminates in a paradoxical and a peculiar situation as it points to the co-existence of absolute and relative.

A meta-analysis of Kuhn's arguments culminates in a situation that is not different. One of the chief themes of his thêses is that paradigms guide research in terms of observation and interpretation of data. If his premise is true – he has, of course supported it with a lot of evidence – then, by extension it can be said that he has culled out data from the body of the history of science to support his theory that paradigms guide research. In other words, the data was collected with the theory – paradigm guides research – already in his mind. As soon as we recognize this, Kuhn's arguments turn on themselves, thus assuming circularity. *A paradoxical situation emerges again: Kuhn's arguments are true and false at the same time.* True because there is evidence to support his claim and false because he contradicts himself by inviting his arguments on himself (alternatively, his arguments have been designated as self-referential by his critics, and have been termed as self-refuting).

Secondly, Kuhn has cited evidence to show that facts and data have no meaning in themselves; they acquire meaning when interpreted against a theory or framework. There is an implicit circularity and paradox here too. By force of Kuhn's arguments, it can be argued that the evidence that he has shown to demonstrate the truth of his arguments is meaningful only against his contention that evidence has no meaning in the absence of a framework. Evidence lends support to his theory whereas the similar kind of contradiction mentioned above, and the fact of being oblivious to his own subjectivity while attributing the crucial role of the scientist's subjectivity in guiding research, renders Kuhn's theory problematic. If the evidences of the other scientists are not sacrosanct, it can as well be said that Kuhn's are not either.

Does this mean that these paradoxes are the final truth about

our existence? Does this mean that they cannot be resolved? Does it mean that we have reached the final summit of our knowledge pursuit? The answer is a resounding no when we begin to explore the Madhyamaka school of Mahayana Buddhism. The philosophy of this school is called *Sunyavada* or *Sunyata*; in the West it has been named the doctrine of emptiness or openness.

Sunyata *and the resolution of the paradoxes*

The simultaneous existence of right and wrong, true and false, and relative and absolute that produces a paradox for Western philosophers poses no problem for the Madhyamika philosophers. In this section, we will examine how these paradoxes can be resolved paving the way for enlightenment or intuitive knowing. The philosophy also addresses the aforementioned issue of Self-Referentiality (which has been used by the critics to refute Kuhn) since a similar kind of charge had been leveled against it, as its opponents found that the discourse after deconstructing all epistemological and ontological presuppositions devours itself without leaving any trace. As we will see, the criticism does not bother the Madhyamika philosophers as they regard the issue of Self-Referentiality as a very valuable tool for the purpose of liberation. This will be commented upon later in this paper in some detail.

Nagarjuna, a second century Madhyamika philosopher, stated that concepts, events and entities do not exist in isolation but exist in relation to one another. He further contends that concepts, events and entities that he calls *swabhava* lack any intrinsic existence, and any attempt to reduce them to having an independent status will lead to absurdity. According to McCagney (1997), Nagarjuna defines *swabhava* as "*nirapreksha paratra* (independent of others), *ahetu pratyaya* (without cause or conditions), *nitya* (permanent or unchanging), *aparijnana* (unknowable), and *akrirayate* (unmade)" (p. 61). Nothing exists in-itself and of-itself, and no concept has any meaning independent of a relation. This is the principle of *pratitya samutpada* or dependent origination, and the main philosophy under which this is discussed is called *sunyavada* or *sunyata* or the doctrine of emptiness. There is no good without evil, no black without white, no valleys without mountains, and no friends without enemies. Darkness is born out of light and day is born out of night. True contains false and right encompasses wrong. Nagarjuna writes:

> How, indeed, will disappearance exist at all without origination? [How could there be] death without birth? There is no disappearance without [prior] origination.
> It does not obtain that origination and disappearance are the same thing.
> It does not obtain that origination and disappearance are different. [Cited in McCagney 1997: 59]

McCagney (1997) citing from the Nagarjuna's work *Shunya-saptatikarika* writes:

> Without one [*eka*] there are not many [*aneka*]. Without many [*aneka*] one [*eka*] is not possible.
> The father is not the son, the son is not the father. Neither exists without being correlative.... (p. 60)

Employing the principle of *pratitya samutpada* or dependent origination and his dialectical skills, he refutes the contention of the realists that a thing exists in-itself or of-itself. The subject does not exist independent of the object; neither does the cause exist without the effect. These dichotomous pairs – like all that we use in everyday life – have no meaning beyond their relationship with one another.

Thus, the rigid dichotomy between the subject and the object crumbles down, for the Mādhyamika critique shows that the act of knowing is a product of the interaction between the observer/knower and the observed/known. Any dichotomized way of thinking results in *avidya*, loosely translated as spiritual ignorance, which motivates the mind to grasp thoughts as things to be grasped by the individual ego. The solution to the enigma of our existence or the knowledge of the ultimate is gained by the transcendence of all the reified and rigid thoughts through a way of 'seeing' and 'being' called *prajna*. The search for knowledge is grounded in our language, pre-suppositions and all those concepts and entities that we hold on to as givens in our everyday life. We attach transcendental and eternal value to these givens, which Mādhyamika deconstructs by placing them in a sociolinguistic and historical context, thus paving the way for a spiritual seeker to transcend the rationalistic tendency to make sense of the truth through any epistemological or ontological suppositions. Huntington (1989) explains this most beautifully:

According to the Mādhyamika, a . . . convoluted and subtle relationship holds between any two dichotomies of conceptual thought, whether expressed in ontological, epistemological, ethical or any other terms: Cause/effect, subject/object, substrate/predicate, absolute/relative, truth/error, good/evil, and all other dualistic concepts find their meaning in the context of their elusive relationship with each other and with an interrelated network of other such concepts. The structure that they give to all experience—a structure that seems "to emerge from the things themselves"—is also dependent on an illusion similar to the Necker cube where each image finds its meaning and existence only in the context of its relationship to partners that must always remain out of sight. The critical difference is only that the context of everyday life in which these other relationship are embedded is infinitely more complex, for it embodies an indeterminate number of historical and circumstantial factors shared by the sociolinguistic community in which this vocabulary is used and thought and perception take place. (p. 121)

Does this mean that the doctrine of *sunyavada* is the absolute truth? The Mādhyamika philosophers would have answered that this question itself was redundant for them, for to answer in affirmative or negative would be to play the game that they were trying to deconstruct. They did not take any position with regards to any epistemological or ontological issues, after having deconstructed all the claims of their opponents. It had been pointed out by the opponents of Sunyavada that the philosophy was incapable of defeating the claims of other philosophies, as it invited the application of its own arguments on itself (a critique of Self-Referentiality not dissimilar from what Kuhn and some deconstructivist and relativist philosophers are facing today). Chandrakirti, a notable seventh century A.D. Mādhyamika philosopher, who is said to have consolidated the *prasangika* branch of Mahāyāna Buddhism, while responding to one such accusation states:

The problem of connection between argument and counter-argument is only a problem for those who presuppose some form of absolute, as you do, and are therefore compelled to meet your claims with appropriate counterclaims. For us, it is a pseudo-problem, because we do not hold such presuppositions. Our words are like the reflection of a face in the mirror – there is no real connection between the reflected image and the face, but the image nevertheless serves a specific purpose for the person using the mirror. Similarly, our words bear no intrinsic connection with your epistemological and ontological problems and the language used to express these problems, but nevertheless these words of ours can serve to realize a specific purpose: They can be understood to express something that is not at all susceptible to expression in the language of "objective facts." [Cited in Huntington, 1989, p. 54]

Similarly, Nagarjuna stated that he had no proposition (*pratijna*), and hence he had no fallacy. In addition, he warned against the dangers of making *sunyavada* an absolute truth by stating that whoever tried to make this treatment a *drishti* or a philosophical point of view was indeed lost. *Sunyavada,* according to Nagarjuna and Chandrakirti, had a purpose, and that was to show the limitation of logic and the discursive mind for the attainment of *prajna,* and for a way of living totally devoid of antipathy, clinging and delusion (when I say the attainment of *prajna,* I do not mean to imply that there is a reality 'out there' that needs to be discovered). Use of logic for any purpose other than to undercut logic can become a dangerous snare, was the warning of Chandrakirti. Clinging to emptiness would be like the water catching fire that was used to dowse fire. The purpose of *sunyavada* is to end fear and suffering for all sentient beings, and to end clinging to every concept of 'truth' and 'reality' including the doctrine of emptiness. When contested by fellow Buddhists that the doctrine of emptiness rejected the four noble truths and the three jewels that Buddha taught, Chandrakirti and Nagarjuna responded by saying that their opponents misunderstood the doctrine because they misunderstood the purpose. Writes Huntington (1989):

Its purpose, as stated by Candrakirti, is to eradicate the inner tendency of conceptual thought to construct reified notions of being (*bhava*) or non-being (*abhava*). Such reified notions generate the philosophical positions referred to as absolutism and nihilism. Even more crucial, though, from the Buddhist perspective, is Candrakirti's point that both of these theoretical positions are representative of exactly the sort of conceptual diffusion (*prapanca*) that lies at the root of clinging and antipathy

> and therefore all forms of fear and suffering, This idea of purpose or application is the pivot on which Candrakirti's philosophy and soteriology turn. (p. 30)

Does this mean that the philosophy had only soteriological aims, and had no empirical basis? The answer is not simple, for a Madhyamaka philosopher will not give a simple yes or no answer. Indeed, Nagarjuna said: "Everything is real, not real, both real and not real, and neither real nor not real: This is the teaching of the Buddha." (Huntington, 1989). Since academic writing is still embedded in the dualistic exercise of proving and disproving things, may it suffice to say that Buddha had the experiential insight of *pratitya samutpada* just before his release from *samsara* (McCagney, 1997). At the same time, the Buddha while explaining *samyak jnana*, mistranslated as the right view, said to Katyayana:

> Katyayana, everyday experience relies on the duality of "it is" and "it is not." But for one who relies on the Dharma and on wisdom, and thereby directly perceives how the things of the world arise and pass away, for him, there is no "it is" and no "it is not." "Everything exists" is simply one extreme, Katyayana, and "nothing exists" is the other extreme. The Tathagata relies on neither of these two extremes, Katyayana; he teaches the Dharma as a Middle Way. [Cited in Huntington, 1989, p. 37]

Not only Kuhn but also thinkers like Rorty, Derrida, Foucault and Gadamer (the paucity of space precludes me from going into the details of their philosophy) have thrown serious epistemological and ontological challenges that are obliterating the rigid dichotomy between subject and object. We have had various responses to their critique – some have qualified them as nihilist, some self-refuting etc. But as we have seen, the matter is more serious than merely taking an either/or stance about such issues. Mādhyamika philosophy not only provides us with a radical way of perceiving the situation, but also bails us out of an impasse generated by the crumbling down of the subject/object and various other dichotomies. Though at a very slow pace, it appears to me that as a civilization we are moving towards the realization of the 'truth' of the 'highest' meaning – *parmārtha satya* – or the *advayajnana* from which will result the *prajna*. It is the claim of the Indian mystics that *prajna*, translated as wisdom, will give us an understanding of the workings of mind and consciousness, if it

can be safely assumed that the problem of mind and consciousness is intricately entwined with the questions of nature of reality and Truth.

Beyond Mind: The appropriate methodology of mind and consciousness research

As it has been pointed out before, the objective of Madhyamika philosophers was to show the limitation of logic and discursive mind as tools for penetrating into the ultimate secrets of our existence. Mystics consider mind—with reason, logic and intellect as its instruments – an inferior and inadequate instrument for knowing the mysteries that surround our existence. This is a consistent thought in almost all spiritual traditions. According to Sri Aurobindo – one of the greatest mystics of the present time – mind and its instruments cannot perceive the Reality as a whole as its very nature is to classify, discriminate, categorize, divide, compare and measure. It tries to understand things through categories, concepts and formulas. In the intellectual history of mankind, there have been scores of such formulas and theories, but nothing definitive can be said about the fundamentals of our existence, despite the fact that most theories have almost equal intellectual appeal, and have evidence to support their claim even if they contradict each other. Sri Aurobindo (1958) puts it most succinctly:

> [I]ntellect is incapable of knowing the supreme Truth; it can only range about seeking for Truth, and catching fragmentary representations of it, not the thing itself, and trying to piece them together. Mind cannot arrive at Truth; it can only make some constructed figure that tries to represent it or a combination of figures. . . . There have been hundreds of these systems and formulas and there can be hundreds more, but none can be definitive. Each may have its value for the mind, and different systems with their contrary conclusions can have an equal appeal to intelligences of equal power and competence. All this labour of speculation has its utility in training the human mind and helping to keep before it the idea of Something beyond and Ultimate towards which it must turn. But the intellectual Reason can only point vaguely or feel gropingly towards it or try to indicate partial and even conflicting aspects of its manifestation here; it cannot enter into and know it. As long as we remain in the domain

of the intellect only, an impartial pondering over all that has been thought and sought after, a constant throwing up of ideas, of all the possible ideas, and the formation of this or that philosophical belief, opinion or conclusion is all that can be done. . . . If the intellect is our highest possible instrument and there is no other means of arriving at supraphysical Truth, then a wise and large Agnosticism must be our ultimate attitude. Things in the manifestation may be known to some degree, but the Supreme and all that is beyond the Mind must remain forever unknowable. (pp. 169-70)

Reason, which Western philosophy has boasted as the panacea of ills and as a solver of all mysteries has failed to deliver, for there is no one Reason. The reason of individuals varies according to their belief, upbringing, attitude, culture, language and perspective. That reason which has the power and money to back its claim comes to be defined as the right reason. Recognizing the relativity of reason, Sri Aurobindo (1958) states:

> You believe according to your faith, which is quite natural, he believes according to his opinion, which is natural also, but no better so far as the likelihood of getting at the true truth of things is in question. His opinion is according to his reason. . . . How is reasoning to show which is right? The opposing parties can argue till they are blue in their face – they won't be anywhere nearer a decision. . . . But who can look at the world as it is and say that the trend of things is always (or ever) according to the right reason – whatever this thing called the right reason may be? As a matter of fact there is no universal infallible reason which can decide and be the umpire between conflicting opinions; there is only my reason, your reason, X's reason, Y's reason multiplied up to a discordant innumerable. Each reasons according to his view of things, his opinion, that is his mental constitution and mental preference. (p. 178)

As it has been stated before, Kuhn's critique has totally destroyed the notion that any kind of objectivity can be expected out of science with regards to fundamental truths. The basic reason this has happened is because the domain of scientific research has mainly been intellectual. To elucidate this point more clearly, let us take a journey back into the first few pages of this paper where Kuhn's ideas are discussed. As we previously noted, the education of scientists

in a paradigm through *disciplinary matrix* and *shared exemplars* orients them to approach the problems in a similar way, so that they do not work at cross-purposes. It gives them a framework to proceed with their puzzle-solving activity, and interpret their data. A change in paradigm results in viewing the same data in a different way; the world begins to appear to a scientist differently than before. As has been pointed out by Kuhn, the history of science reveals that there are many competing theories before one paradigm becomes dominant, and many different explanations are possible from the same set of data. The fact that various meanings can be ascribed to the same data is because of the difference in constitution of the minds that approach it. Sri Aurobindo (1958), who had anticipated Kuhn in his main thesis, explains that this happens because of the fundamental nature of the mind:

> Objective external physical things are seen very much in the same way by human beings because of the construction of the mind and senses; with another construction of mind and senses quite another account of the physical world would be given – Science itself has made that very clear. (p. 205)

Thus, we see that any enterprise for mind and consciousness research that is embedded in a subject-object dichotomy is doomed for failure. The mind – with reason, intellect and logic as its servants – too cannot know the secrets of ultimate reality. This brings us to the discussion of the appropriate methodology for the inquiry into mind and consciousness, and the ultimate truth of our existence. It is through the transcendence of mind, logic, intellect and reason that the mystery of mind and consciousness can be solved. Mind can only take us so far as to show us its limitation, as the previous few pages may have shown. It can show us how ignorant we essentially are with regards to the basic enigma of our existence: Where do we come from and where do we go? Using mind in order to solve the problem of mind and consciousness is like trying to look at one's own face without a mirror. The mystics claim that the deeper truths of our existence unravel themselves on a silent mind, compared metaphorically to an ocean that is absolutely calm. In other words, stillness of the mind is the necessary condition for accessing knowledge that lies beyond the domain of intellect. It is this region that holds the key to the secrets of mind and consciousness. Sri Aurobindo (1958) states that the "pure stillness of mind is always

the required condition, the desideratum, but to bring it about there are more ways than one" (p.193). A complete silence of the mind and a change of ordinary human consciousness hold the promise of accessing knowledge of the fields not available to the physical eye. Writes Sri Aurobindo (1958):

> The mind can think and doubt and question and accept and withdraw its acceptance, make formations and unmake them, pass decisions and revoke them, judging always on the surface and by surface indications and therefore never coming to any deep and firm experience of Truth, but by itself it can do no more. There are only three ways by which it can make itself a channel or instrument of Truth. Either it must fall silent in the Self and give room for a wider and greater consciousness; or it must make itself passive to an inner Light and allow that Light to use it as a means of expression; or else, it must itself change from the questioning intellectual superficial mind it now is to an intuitive intelligence, a mind of vision fit for the direct perception of the divine Truth. (p. 174)

A change of ordinary human consciousness becomes a necessary condition for mind and consciousness research. It is not by looking outside of us that we can find answers to the enigmas that shroud us, but by looking within. In this research the researcher and the researched become one; the subject becomes the object and vice versa, leading to the transcendence of the strict dichotomy of subject-object that Western science has practiced so far. Yoga, which actually means a union with the Divine or with the essential ground of all beings – whichever way one may want to see it according to one's preference – is the key through which a change of human consciousness – and hence a transcendence of mind – is possible. Therefore, it is the appropriate methodology of mind and consciousness research. The Indian mystics have practiced this art for centuries together, and have left behind a rich source of literature for all kinds of aspirants who want to take this path. The mystics have always stated that there is no one right way to take; that is the reason why there is a plethora of paths leading to the oneness that underlies this Universe, based on the different constitutions and psychological make-up that humans have. If the West has perfected the art of researching the material universe, it would not be an exaggeration to say that the Indian mystics have explored the area of consciousness

inside out. However, when I talk about the distinctiveness of India in this field, I do not mean to suggest that such experiments have not been conducted anywhere else in the world – what I definitely mean is that they have been fewer in other parts of the world limiting the many possibilities of approaching the consciousness that humans have in their repertoire.

Concluding remarks

This brings me to comment on one of the aims of this conference, which is to explore whether Indian theoretical traditions can offer useful insights and solutions in the field of mind and consciousness research. If Kuhn's critique of science, emanating from his analysis of its history, has pulled the rug from the feet of the scientists, Indian philosophy and mysticism should give them hope of finding the ultimate secrets of our existence as I have discussed in this paper. However, I would like to object to the use of the term theoretical for the Indian traditions. These traditions have not come out of speculation or out of superstition, as the dominant West would like us to believe, or has made us believe by systematically destroying the Indian education system under its colonial rule. The parallel for the word philosophy in Indian literature is *darshana*, which suggests that the thoughts in India do not owe their genesis to speculation and mental reasoning as has happened in the West, but to experience and seeing. The yogis, *rishis* and *munis* experienced the subtle realms through the various yogic practices, and then either recorded their experiences or had them passed on to the next generation through the oral tradition that was prevalent in Ancient India. Despite the fact that the tradition of yoga has suffered a setback, great yogis have consistently graced the earth of India. Sri Ramakrishna, Sri Aurobindo, Raman Maharshi and Devraha Baba are recent examples. Yoga has a law of its own, and it deals with very concrete realities that can be seen with a non-ordinary vision, known as the third eye in the parlance of the Indian mystics; it is not a compendium of beliefs held by savages or barbaric people. Sri Aurobindo (1958) explains this most succinctly:

> The mystic goes beyond into a region where this mental basis falls away, where these data are exceeded, where there is another law and canon of perception and knowledge. His entire business is to break through these borders into another consciousness

> which looks at things in a different way and though this new consciousness may include the data of the ordinary external intelligence it cannot be limited by them or bind itself to see from the intellectual standpoint or in accordance with its way of conceiving, reasoning, established interpretation of experience. (p.197)

And again,

> [T]he experiences of yoga belong to an inner domain and go according to a law of their own, have their own method of perception, criteria and all the rest of it which are neither those of the domain of the physical senses nor of the domain of rational or scientific inquiry. (p. 206)

In the formative stages of the paper, I have quoted extensively from the writings of Kuhn as must be apparent. The reason was not only to show how the rationalistic scientific research has culminated in relativism, but also to expose the various assumptions that underline a scientific research that are not put to any kind of 'scientific' test. A scientific research before it proceeds defines what is scientific, what is not, and what is admissible in the canon of scientific beliefs and what is not. Science has its own body of beliefs and superstitions that have not been adequately researched, critically examined, or been put under the litmus test of 'scientific' inquiry. Rationality and objectivity are grand theories that have been constructed by science – which basically is the progeny of an imperialistic culture – whose main aim was to control, dominate and plunder. The science of yesteryear, without exploring the field of yoga on its own terms and the methodology that it uses, with one sweep put it in the dustbin of superstition, thereby obliterating the wisdom of a civilization which is one of the oldest in the history of civilizations – a civilization which had influenced practically the entire world with its ideas without using force. Sri Aurobindo (1958) states this most emphatically:

> It is quite true that the word "superstition" has been habitually used as a convenient club to beat down any belief that does not agree with the ideas of the materialistic reason, that is to say, the physical mind dealing with the apparent law of physical process and seeing no further. It has also been used to dismiss ideas and beliefs not in agreement with one's own idea of what is the rational norm of supraphysical truths as well. For many ages man cherished beliefs that implied a force behind which acted on principles unknown to the physical mind and beyond the witness of the outward reason and the senses. Science came in with a method of knowledge which extended the evidence of this outer field of consciousness, and thought that by this method all existence would become explicable. It swept away at once without examination all the ancient beliefs as so many "superstitions" – true, half true or false, all went into the dustbin in one impartial sweep, because they did not rely on the method of physical Science and lay outside its data or were or seemed incompatible with its standpoint. (pp. 246-7)

Thus, the Indian yogic tradition has a lot to offer to the world in matters of mind and consciousness research. At the same time, it is time for India to become more open to the traditional wisdom of her ancestors – the *rishis, munis,* and the yogis – and research mind and consciousness the way it has been laid down by them– and there is no one way. There are many from which one can choose according to one's constitution and preference. It is time for this ancient civilization to put a stop to getting evaluated by the "rationalistic" propaganda of the West, and put more faith in the countless sages that it has witnessed, so that it is once again able to recreate the great civilization that it was in the past, and help the world to become a more peaceful and equitable place to live in.

REFERENCES

Aurobindo, S. (1958). *On Yoga II: Tome One*. Pondicherry: Sri Aurobindo Ashram Press.

Capra, F. (1992). *The Tao of Physics* (3rd ed.). London: Flamingo.

Chalmers, A. F. (1982). *What is This Thing called Science?* Queensland: University of Queensland Press.

Feyerabend, P. (1993). *Against Method* (3rd ed.). New York: Verso.

Huntington, C.W. (1989). *The Emptiness of Emptiness*. Honolulu: University of Hawai Press.

Kuhn, T. S. (1970). *The Structure of Scientific Revolutions* (2nd ed.). Chicago: University of Chicago Press.

Leahey, T. H. (1991). *A History of Modern Psychology*. New Jersey: Prentice Hall Inc.

McCagney, N. (1997). *Nagarjuna and the Philosophy of Openness*. Boulder: Rowman & Littlefield Pblishers, Inc.

Weber, M. (1946). Science as a Vocation. In H.H. Gerth & C. Wright Mills (Eds. and trans.), *Max Weber: Essays in Sociology*. New York: Oxford University Press.

Whorf, B.L. (1962). *Language, Thought and Reality: Selected Writings of Benjamin Lee Whorf*. Cambridge, Mass.: MIT Press.

Person and Boundedness in Wittgenstein and Tagore

Positioning Artificial Intelligence

ABRAHIM H. KHAN

At times of development in civilization, whether in culture or technology, there are prevailing concepts that influence our understandings. Descartes' concept of the mind and its contents became a starting point for classical psychology, Locke gave us the idea of sensory immediacy as a criterion for the real, and Hobbes' suggestion of building complex ideas from simple ones resulted in the genetic method.[1] The twentieth century is certainly no different, with respect to philosophy, science, and technology. For example, many classical problems in philosophy and psychology were transformed by a new notion of mechanism due largely to the work of Alonzo Church and of Alan Turing. A different level of analysis emerged from the Church-Turing thesis, independent of research directions in physics, but still mechanistic in spirit. It led to the formation of the field of computer science, and more specifically Artificial Intelligence (AI). One of the concepts gaining ascendancy in this field is that of person or personhood.

The concept "person" has central to its understanding and contours a particular feature that must be considered in any adequate account of it. To focus on that feature, I attend to the semantic field of the term 'person' and then to writings by Wittgenstein and Tagore to sharpen our understanding of it. In their wrestling with language these two individuals sought to convey through language its boundary and that which is on the other side of the boundary. The logic of language is partly bound up with the concept of person and has implications for research developments relating to biogenetic engineering, organ transplants, and AI. Since a research line in the latter area is tending to entertain claims relating to the person, it is all the more pressing that cognizance be taken of the feature so

that AI might position itself properly in its attempt to translate or re-interpret the concept. Some concepts are like individuals in that they do not withstand the ravages of time yet retain a homesickness for scenes of their childhood.[2] Personhood is one such concept.

Among some researchers in AI and cognitive psychology is a growing feeling that human intellection cannot be understood along traditional lines of sensation, perception, memory, comprehension, reasoning and so on. The boundaries of these categories will have to be redrawn to properly account for the integration of cognitive activity.[3] While there is no agreed common understanding of the notion of intelligence among the federation of disciplines known as cognitive science, intelligence seems to have at least four components: 1) systematically structuring elements of knowledge and performing logical operations in complex systems, 2) intuition and creativity, 3) social empathy, and 4) the integration of the rational and the emotional with the sensory and motoric actions.[4]

Human intelligence is also inextricably linked to self-consciousness or the capacity to reflect on one's own thinking or to be aware of one's mental acts. Computational machines do not as yet possess this capacity, which is a necessary prerequisite for controlling creatively one's own line of thinking, or mental acts such as choosing alternatives. However, for some computer science advocates, intelligence and the capacity to reflect are aspects of the human person, which is a conceptual idea best understood as a physical object. In 1981, John Haugeland explained that the idea that is basic to cognitive science is that intelligent beings are semantic machines, and that people and intelligent computers turn out to be merely different manifestations of the same underlying phenomenon.[5] Eight years later, John L. Pollock contended in his book, *How To Build a Person*, that people are physical objects that supervene on their bodies but are not identical to their bodies, in the manner of a statue consisting of or supervening on a lump of clay.[6] Pollock goes on to say that there is no such thing as the concept of a person, that what we mean by that idea is vague, and that it is our task to construct a more precise concept. My proposal is that a more precise concept would have at the least some minimal understanding of what is meant by the idea of person in the humanities, and would reflect some aspects of its childhood scene.

The concept of person is at the heart of my presentation. I take this concept as one of the dominant ones to constitute a challenge

for AI. Headway in the field of computer science would require also taking this concept seriously, by thinking about and with it. Here, I draw attention to an aspect of it that is overlooked and would constitute a *hard problem* (to use a terminology from David J. Chalmers who distinguishes between the easy and hard problem in consciousness studies).[7] Especially if machines are to be truly intelligent or to be considered as persons, it is important that researchers and theoreticians in the field understand the scope and limits of this particular concept. This is all the more important to ensure that the concept is not made subservient to technology, and that its employment is in a fruitful rather than retrogressive way.

Recent literature on the nature and enterprise of cognitive science seems to be suggesting a direction which is dubbed strong psychological AI in contrast to one dubbed weak AI. Accordingly, the computer is understood as a mind, a perspective that anticipates a vision of it having civil rights, carrying on romances with humans, and so on. There is already talk of affective computing,[8] which Hollywood makes good through its Star Trek creations of Data on the Enterprise and of the holographic Doc on Voyager. Even bolder is the research direction dubbed supra-psychological AI. Its proponents exploit "the generality of the term 'intelligence' and suggest that there are principled reasons for extending its meaning to new domains."[9] In this accounting there is nothing logically incoherent in the idea of an artificial intelligence that duplicates human intelligence or about conceiving the brain as having synthetic neurons, and thereby being prosthetic brains to the same extent as prosthetic lungs and hearts. The similarity in the mental life of artificial intelligence and that of human intelligence is in functional organization. This form of functionalism as a solution to the mind-brain problem has for its critic John Searle whose objection to the Turing test is that it is a reverse discrimination against humans.[10] That is, humans behave as if they have not only intentionality, consciousness and free will, but also the right sort of private experience and bodies as well. The point of the objection is that mental types are not necessarily reducible to physical types.

Properly understood, the concept of person illustrates further the difficulty in reducing a mental type to a physical type, and hence the challenge that AI has to address in order to apply the term person coherently to thinking machines or prosthetic brains. Merely imputing intelligence to computing machines does not make a machine-being into a person, for the criteria used in the imputation are ones that are based on biological, behaviourial, and phenomenological evidence. The concept of person has also a dimension or framework that is metaphysical, and to which we are constantly striving to give expression. That striving has partly to do with the experience of boundedness. Becoming a person is tied up with our experience of a limit or boundary in our thinking and with our striving to express that which is on both sides of the limit. This striving may in a way represent the Frame Problem confronting strong AI.

Personness has more than a material and social context. Its definition includes the idea of boundedness. Human beings are distinguished from other life forms by virtue of encountering a limit in their thinking and having a longing to express not just that which is on one side of the boundary/limit but also that which is beyond it. This longing may express itself in terms of a silence (as with Wittgenstein, Kierkegaard, and the Buddha) that has a survival value, and hence a crucial factor in the workings of mind and consciousness. My contention is that our notion of being fully human or truly a person, carries with it the idea of boundedness, or the potential for experiencing it. To understand this feature it is important to distinguish between the individual human and the person. While every individual is a human to whom respect and dignity are due from a liberalism of rights perspective, not every individual is necessarily a full person as yet. The reason is that person is not a status concept but a self-involving or actualizing one. It is connected with the phenomenon of boundedness or longing to transcend that which limits or restricts, to get behind the limit or boundedness that is experienced.

To examine closer this feature, let us look at the meaning complex for the term "person". The term is derived from the Latin *persona*, from which semantically subtends two ideas that form its core meaning. One of the ideas is suggested by *personare*, meaning to sound through, as in the case of the voice of an actor resounding (*sonare*) through (*per*) a mask. Philologists think that the word for mask might be *phersu*, which is perhaps of Etruscan origin or borrowed from the Greek *proposon* that signifies primarily mask, and secondarily role. Either way, whether it be mask or role, *persona* came to suggest the notion of a role (personage) or type or character superimposed on an individual. Among Greek and Latin moralists

the meaning of *persona* came to take on a moral tone, a sense of being conscious, free, and responsible. From this extended sense it was a short step to the juridical meaning of *persona* as an individual with legal and moral rights. This is one of the meaning components that is entrenched in daily speech when person is used. What is important to observe in this aspect of the meaning complex derived from *persona* is the phenomenon of limit. The mask or role limits its wearer/actor who as surplus longs to express more than the part or role assigned. It limits also what is seen by the audience and creates in the audience a longing to transcend or get behind the mask, to reach out as it were to the real individual – the one acting or wearing the mask.

The other idea that the Latin *persona* subtends is that of *personnalite/* substance that is human and even divine. It is characterized by the idea of tearing away superimposed layers, to lay bear the nature of the role player, or that what is one in and of itself (*per se una*). Its ambiguity in reference (human or divine substance) yields the idea of human substance as being open to the possibility of divinity, as going beyond its limits. This idea gained prominence in the fourth and fifth century theological controversies in Christendom over persons in the Trinity. Boethius was to later add on the idea of rationality to the formation of an indivisible and whole substance to yield our classical definition of person. He was rendering in Latin terms that which is expressed in Greek by Neo-Platonism. He provided a metaphysical foundation for the notion of person. On that foundation Christianity developed a philosophy of personality. It was then borrowed and altered by modern philosophy. Peter Strawson, for example designated person as a primitive concept associated with predicates, some of which are attributable to material bodies,[11] and seem to dominate in our quotidian conception of person.

In fact, one of the accomplishments of modern philosophy is that it defined the concept of personality in accordance with psychological knowledge. Its reason for doing this was to preserve humanity's distinctive position in the face of tendencies to speak of a general uniform order in the world. Descartes, for example, emphasized human consciousness, Leibnitz placed the true essence of human personality in self-consciousness, and Kant deepened the ethical view of personality by defining it in terms of freedom and independence from the mechanism of the whole of nature. In Fichte the notion underwent further transformation to become the category of self *(moi)*, which is already a primordial category in the Pietist

tradition, but becomes a central category in contemporary philosophy.

The social sciences followed the lead of modern philosophy by redefining personality in terms of observable behaviour and emotional tendencies. That is, personality came to have for its reference a socially perceived individual, or the organized stimulus and response characteristics of an individual dynamically involved in social situations. Hence, the concept personality came to imply an inference from behaviour. It is this sense of the concept that has gained ascendancy in ordinary language,[12] and is predominant in cognitive science discourse. Such limited meaning does not tally with, is far removed from, the idea of personhood understood as the opening up of oneself to transcendence in order to become fully a self. It misses the phenomenon of boundedness that is resonant in the root meaning of person, and associated with reaching beyond our limits to become intimately related. The use of the adjectival form "personal", in daily speech underscores the intimacy that is integral to the concept person.

Indian thought mirrors this phenomenon of person by a different set of concepts: *purusa, prakṛti,* and *jivapurusa.* The first two form the starting point in the Sankhya philosophical school and its emphasis on evolution. The name of this school or system of thought is etymologically associated with the idea of exact knowledge or number,[13] and hence discrimination or observing boundary. Integral to understanding of evolution are notions of boundednesss and of the reaching over boundary. *Purusa,* designating the primordial person and a male principle, corresponds to pure consciousness. It provides the initiative through the medium of consciousness for the evolving of the physical world of *prakṛti.* By itself *prakṛti* is formless, undifferentiated, even though it is dynamic matter – the female principle. The interaction between the two, which is only a semblance of contact, occasions creation, including the *jivapurusa* or phenomenological self.

According to the Sankhya interpretation of reality, both *prakṛti* and its product *jivapurusa* are bounded, and must reach over their boundedness to become free. In the case of the *jivapurusa,* its reaches over by desisting to identify itself with *prakṛti,* by realizing itself as purely *purusa* through the discrimination (*viveka*) of *purusa* from *prakṛti.* Sankhya proceeds from these foundational ideas to its analysis of the question "Who am I?", observing distinctions among reason

(*buddhi*), ego (*ahamkara*), mind (*manas*), and consciousness (*cit*), all of which are necessary in its explanation of the stuff of the universe. The exposition of each of these attributes are not required here to draw attention to a theme they sustain and that characterizes much of Indian thought: boundedness and the reaching over the boundary to become related simultaneously to what is on the other side of it. This theme is reflected also in the thought of some modern Indian thinkers such as Aurobindo and Rabindranath Tagore.

Contemporary western thought counts Kierkegaard as one of its resources for thinking about the theme. He presents the self as striving to become itself by a relationship in which it is related finitely to the finite, and at the same time infinitely to the infinite. His analysis of becoming a true self has an impact on reflections by later thinkers such as Heidegger and Levinas. Wittgenstein is another who drew insights from Kierkegaard as well as from Tagore,[14] each of whom is concerned also with the working of language in depicting what it means to be a genuine human person. Kierkegaard distinguished between words used ordinarily by a sensate-psychical individual, by one who knows himself according to the conditionals of knowledge governing the mundane world, and those very words used by an individual conscious of human life as more than another item corresponding to the order of things in the world. Though the two are connected by the use of the same words or phrase, there is a world of difference between them. The latter, conscious of himself as having an existence that is more than mundane and speaking about that existence, uses the language metaphorically. That is, he has for Kierkegaard "made a transition or has let himself be led over to the other side; whereas the other has remained on this side."[15]

Tagore, in his reflections on the problem of self, noted that "man's words are not a language at all, but merely a vocal gesture of the dumb. . . . The more vital his thoughts the more have his words to be explained by the context of his life," and that those seeking the meaning get only to the house and "are stopped by the outside wall and find no entrance to the hall."[16] Language clearly is intricately connected in representing person, and is made more evident in both Wittgenstein and Tagore. I turn to these two thinkers in limning that feature of person which constitutes a challenge that AI must address, in properly appropriating the concept of person.

While Wittgenstein offers no positive account of the nature of the self, his writings contain a few sparse but insightful remarks

regarding the notions of the limits, the ethical and the self in relation to the logic of language. In fact the remarks of the *Tractatus* do not readily make a connection between the ethical dimension and logical form. But as he himself noted, the aim of that book is to set a limit to thought. Language is for him the locus for setting that limit with respect to what can and cannot be said. He noted too in a letter to Ludwig von Ficker, that the book's point is an ethical one, drawing limits to the ethical from the inside, thus producing two parts to the work. Or, as he puts it "my work consists of two parts: the one presented here plus all that I have not written."[17] The point to these notations is that Wittgenstein expected in his reader capacities of understanding that are a token of a rich and variegated life, one finely textured, thick with content, and developed to the point that the ethical will be immediately apparent even if it is not overtly expressed.

This is because in his view becoming a person is resident not in any philosophical doctrine or activities, but in a quality of deed, feeling, talking, willing. What it entails then is the conversion of activities into capacities. That is, we look because we possess eyes, – looking is an activity which can be described in non-personal terms. But we all do not necessarily see details or subtle changes in texture, etc., since seeing is a capacity that has to be exercised.[18] As with looking, the same would be true for a number of other activities which are designated by psychologically purposive verbs. The self or "I" is tied up with actions, beliefs and feelings or in short with capacities and not mere activities. Language is a powerful means of becoming a person. In fact personness shows itself through our use of language – judgments, responsible feelings – but the logic of the language will not give to person substance which it has not already acquired.

So the remark in the *Tractatus* that the limits of my language mean the limits of my world would signify that the self is not part of my world but is a presupposition of its existence. Without limits there will be no language or world for me. Logic sets limits to language and to the world, determining what is sensible to say and what is not, but must instead be shown. A fact about the world, about what is happening in it, would be statable as a proposition that makes sense. However, expressions of value may be mistaken for propositions, when in fact they do not tell or explain anything about how things are in the world, or what happens to be the case.

Instead, value expressions are about the other side of the boundary drawn by logic. They treat the world as a whole, transcending the facts of the world, not even mirroring the facts.[19] This is what Wittgenstein meant by *Tractatus* proposition 6.421, that ethics is transcendental. Ethics, as with aesthetics and religion, is concerned with the sense of the world. Value is outside the world and cannot be pictured or expressed in a proposition as something possible. The use of value or aesthetic expression is to get the reader to see the world (an action, or an object) aright, to see that its sense is conveyed.[20] In endeavouring to do so, value expressions go beyond the boundaries of significant language.

Wittgenstein made this point also, in a lecture on ethics, specifically with respect to the miraculous. One of the features of a miracle is that it is inexpressible in the absolute sense and thus is on the other side of the limits of language. To try to say something about absolute value, or the world as a whole, is to attempt to say the unsayable, to go beyond the boundary or logic of language. The point to all of this is that to run into the boundary, to experience boundedness by language regarding what can and cannot be said, does not diminish the tendency or longing to write or talk about religion and ethics.[21] In particular Wittgenstein spoke of ethics as a document of the tendency in the human mind to run against the boundaries of language in trying to say which cannot be said in the manner in which we try to say it.[22] He referred to the boundary in that context as the walls of our cage, and elsewhere as the limits of language, and the net (*Tractatus*, proposition 6.35).

What is important to note, however, is that in speaking about a boundary he was also referring to the phenomenon or experience of boundedness. Each of his metaphors for the boundary indicates that the boundary is more that a logical matter, that it is also experiential and existential as well. Wittgenstein described the longing which is constitutive of person as a tendency to reach out beyond the restrictiveness. One of his conversations relating to Kierkegaard and Heidegger makes this clear. Wittgenstein remarked as recounted by the philosopher Waismann:

> I can well understand what Heidegger means by Being and Angst. Human Beings have a drive to run up against the boundaries of language. Think for example of the astonishment that anything exists. This astonishment can be expressed in the form of a question and also there is no answer at all. Nevertheless, we dash ourselves against the boundaries of language. Kierkegaard has also seen this throwing of oneself and even described it in very similar way (as throwing oneself against a paradox).[23]

The *Tractatus*, on the whole, is also a throwing of oneself, of Wittgenstein wanting to go beyond boundedness which he metaphorically describes as the limits of language. He reaches over to the ethical in the dramatic closure with self-referential propositions that direct attention to the text itself and away from him. He creatively directs his longing to go over the boundedness that language imposes on him by the very use of language to declare that the propositions are elucidatory in this way: (1) whoever understands them would recognize them as nonsense when he has used them to climb up and over them. And (2) the propositions must be transcended in order to see the world aright. The *Tractatus*, more than being a solution to the central problems of philosophy, is overlaid with significance. For it is itself a type of speech that shows what it cannot say, and in this sense it has a mythic dimension – providing a context for universal human meanings, purporting to tell how the world is, symbolizing itself through inversion, and standing in a complex relation to reality. It is the archetypal cave from which one cannot get out unless one has first entered. It is a picturing of what is on the other side of the limits to my world.

For Rabindranath Tagore, boundary/*sema* is an important term in his universe of discourse. Boundedness and longing are pictured through his dramatic play *Raja* or *King of the Dark Chamber*. As with the *Tractatus*, the play is attempting to treat the limits of language through the use of language to depict the longing or capacities that are constitutive of person. The play alludes to the capacities through the depiction of actions, beliefs and feelings by which we become person or genuine self. In brief, Queen Sudarshana longs to see the face of the King. His devotees know him to be incomparable to other mortals, except for the Queen in the early part of the drama who is longing to see him but never does because she meets him only in a dark chamber. Innumerable doubts, difficulties, and misunderstandings bring her to despair and to discuss with her maid who has mastered her own doubts and rage, and who knows the way in and out of the chamber, including the King's movements, without ever having actually seen him. By the end of the drama Sudarshana too succeeds in reaching over her boundedness and realizes her

longing.

The King symbolizes the other side of the boundary, or more accurately the everywhere and nowhere in particular which is not expressed by factual statements. In his prose works, Tagore employs as pointers to what is symbolized expressions such as Universal or Supreme Man, consciousness of the Real, Supreme Unity, or *Purusa* (Being who is Person). He references it further as "the negative idea of the infinite is merely and indefinite enlargement of the limits of things; in fact a perpetual postponement of finitude."[24] Positively described, this aspect of the infinite is the *advaitam*, or absolute unity whose birthing is beyond the revolving finitude of space and time and only "to be realized in one's own inner spirit".

Further, each individual is able to sense it through the faculty of imagination. He describes this faculty as that which "makes us intensely conscious of a life that we must live and that contradicts the biological meaning of the instinct of self-preservation." And in another place he alludes to this capacity in the following way: "man has a feeling that the apparent facts of existence are not final . . . that his supreme welfare depends upon his being able to remain in perfect relationship with some great mystery behind the veil."[25] This feeling or longing to go beyond the facts of our physical existence is the urge to realize completion by a bonding with the *advaitam* and not by denial of our physical reality.

The Queen, however, symbolizes the longing and realization by her becoming the second self of the King. By the end of the drama, the King says to her "that which can be comparable with me lies within yourself."[26] That is, in the mirroring of one by the other, the King tells her that she is his second self. The *jivapurusa* becomes the second self not by merging with *purusa*, but by a relation of itself to itself as part of a seeming interaction of *prakrti* and *purusa*. Tagore reminds us of this fact about proper relation when, in his essay on the artist, he notes that living efficiently requires knowing facts and the laws, but living happily requires one to "establish harmonious relationships with all things," and with whatever we have dealings, for "our creation is the modification of relationships".[27] This is in effect becoming a person through the conversion of activities to capacities.

Boundedness is another aspect that the Queen as *jivapurusa* represents. Pride and misunderstandings, though they may define the phenomenal self, are impediments to the *jivapurusa* properly

relating itself to itself. Out of longing, the Queen falls in love with another king whom she mistakes as her husband King. And in conversation with her maid, she asks: "Why does one make such mistakes?"[28] The maid replies "Mistakes are but preludes to their own destruction." Brought to despair and humiliation, the Queen is ready to become as her maid. That is, she learns to exercise capacities and qualities correlative with right discrimination required for finding the way in and out of the dark chamber, the archetypal cave.

Person from this Indian perspective is constituted by harmonizing within oneself a dual set of desires which through right discrimination are no longer confused. While one set is associated with our finiteness, the other is with an "element of the superfluous in our heart's relationship with the world."[29] This surplus in us shows itself in art which, for Tagore, has for its principal object "the expression of personality, and not of that which is abstract and analytical,"[30] and which "necessarily uses the language of picture and music." While the eye is for the vision of the physical, the faculty of imagination is for the vision of wholeness of personality.[31] And again put differently, he reminded readers that the 'world of our knowledge is enlarged for us through the extension of our information: the world of our personality grows in its area with a large and deeper experience of our personal self in our universe through sympathy and imagination."

Through this latter faculty the impulse to touch the other side of the boundary finds its expression in the production of art, literature and whatever is recognized universally as culture.

The concept "person" retains in Tagore's universe of discourse its etymological mooring by his returning often to the scene of its childhood. That childhood scene is marked by the longing for intimacy with what lies beyond the limit created by our physiology and biology. He calls on the testimony of India's medieval poet-saints and village poets as proof of a direct perception of a longing to actualize what is potential in each individual and thus what defines our humanity. This is the transcendent aspect of ourselves, the formless in the individual forms of humanity, the *nara-narayana* or God-man of the poet saint Rajjab. By God becoming human in each individual, humanity's art and literature are the mask or boundary through which the inwardly real in each individual is manifest or made objectively real. Hunger for reputation, inscribing one's name

on walls of majestic monuments or associating one's name with works of art universally recognized through the ages is, for Tagore, a longing to actualize personhood.[32]

That the human body is more than a physiological mechanism, that it is also divine, comes through Tagore's distinction between creation and construction. He associated science with construction or utility and poetry with creation and self expression, stating that poetry has "to use words . . . which do not merely talk but conjure up pictures and sing. For pictures and songs are not merely facts-they are personal facts. They are not only themselves but ourselves also."[33] He reminded readers too that "the world as a creation is not a construction; it is more than its syntax. It is a poem which we are apt to forget, when by exclusive attention, grammar takes complete possession of our mind."[34] What striking family resemblance to Wittgenstein on language!

Thus far, our referencing here of Wittgenstein and Tagore is to lay out an aspect of the person which has to be considered for the term to be coherently applied to thinking machines or even prosthetic brains. For AI to annex the concept of person, a machine would have to develop a persistence to go beyond what it recognizes to be logically contradictory but finds existentially significant to itself. In Tagore's language, this is the mirroring of the second self of the King, of that which is infinite, and not the self of a finite designer/ creator. Bertrand Russell too may have sensed this aspect of individuality when he exhorted us to free from within ourselves the artist imprisoned in order that he/she would spread joy everywhere.[35]

If, however, some semblance of self-actualizing is reached by a thinking machine, through a closing of the gap between processing powers of machine and the human brain, there would remain the question as to whether a machine with the skills of humans is in fact a person or merely a sentient life form that is useful, and for which Tagore employs the term *shudra* metaphorically.[36] Would it cease to be considered as machine? Aeroplanes soar through the air, fly in formation, and have wings. But they neither flap their wings, nor are considered birds. Does the complex information processing involving feedback loops, pattern recognition, stimulus responses, image projections, redundancy, and show of human skills, amount to a mirroring of a second self or spreading joy to others? Alternatively put, passing the Turing test is simply not enough.

To conclude, from a methodological point of view, Sankhya

thought may prove to be an interdisciplinary resource for meeting the challenge posed, in enriching our understanding of the idea of person, of what it means to be humanly conscious. It has room to accommodate the line of thinking about personness in terms of boundedness and going beyond it as referenced by Wittgenstein and by Tagore and reflected in the meaning complex of the term person. At the same time it accommodates the contention that the absolutes, *prakṛti* and *purusa,* have no real contact (*samyoga*) with each other, but only the semblance of one (*samyogabhāsa*). The Queen as second self of the King is not in direct contact with the latter, but is in a reflexive relation of herself to herself mediated through another (the King). Alternatively put, the potentialities with which *purusa* creates the world belong to *prakrti,* and the realization of self or person is not a separation from the latter.

That being the case, machine lifeforms need not be in direct contact with transcendence or the sacred or absolute. The convergence of artificial neural networking with the physical symbol-manipulating systems opens up possibilities of computers gaining more flexible self-control, by utilizing principles of associative memory, image recognition, and recursivity. In systems taking advantage of that convergence, the recognition of natural symbols, for example, would be enough to create a semblance of contact with that which is apparently absolute, and hence the becoming person. This semblance then approximates personhood and thus holds out a possibility for addressing the challenge that AI faces with respect to ascribing personhood for its thinking machines. Personhood implies striving to reach over an existential boundedness as the discourses of Wittgenstein and Tagore show. For AI to speak coherently about artificial life forms as persons it must work on the hard problem, making the lifeforms reflexive of themselves as being bounded yet striving to reach over their boundedness. And, Sankhya thought may have some insights to offer on that score.

NOTES

1. Susanne K. Langer, *Philosophical Sketches* (Baltimore: Johns Hopkins University Press, 1962), p. 54, draws attention to the phenomenon of exploiting concepts to solve problems conceived in terms of the concepts, and so does Zenon Pylyshyn "Complexity and Study of Artificial and Human Intelligence", in *Mind Design,* ed. John Haugeland (Cambridge, Mass.: MIT Press, 1981), p. 67.

2. Siren Kierkegaard described concepts this way, having in mind those that are pertinent to human existence and had become distorted and transformed beyond recognition. See his dissertation *The Concept of Irony* (Bloomington: Indiana University Press, 1968), p. 68.
3. Zenon Pylyshyn articulates the growing feeling of the inadequacy of traditional categories and examines in the work cited above some ways of redrawing the boundaries.
4. The four are identified by Ulrich Ratsch and Ion-Olimpiu Stamatescu in their introduction to *Intelligence and Artificial Intelligence: An Interdisciplinary Debate*, ed. U. Ratsch, M. M. Richter and I.-O. Stamatescu (Berlin: Springer, 1998), pp. 3-7. The two authors offer a general description of intelligence as the "ability of human beings to react to ever new challenges posed to them by the environment (natural or social) and to solve the problems involved". p. 3.
5. John Haugeland. "Semantic Engines", in *Mind Design* (*op.cit.*), p. 31.
6. John L. Pollock, *How To Build A Person* (Cambridge, Mass.: MIT Press, 1989), pp. 32ff.
His claim rests on a distinction between beliefs/representations that are *de se* and *de re*. He holds to the latter, which is to think "of an object under a description, from that acquire a *de re* representation of the object, and then eventually forget the original description and be able to think of the object under the *de re* representation," p. 23. In the case of *de se* belief (following Roderick Chisholm, David Lewis, and John Perry) the object is a property or concept. The "belief does not take on a propositional object," as in the case of perceptual beliefs, pp. 24f. Pollock then goes on to considerations which seem "to indicate that a person cannot be identical with his body. He only supervenes on his body in the same sense that virtually all physical objects supervene on simpler physical objects.' p. 37. In other words, 'the concept of a person must simply be the concept of a thing having states that can be mapped onto our own.... Obviously on this criterion a machine can be a person." p. 111.
7. David J. Chalmers, "The Puzzle of Consciousness", in *Scientific American* (December 1995) and reprinted in a special issue: *Mysteries of the Mind*, pp. 30-37.
8. Rosalind W. Picard, *Affective Computing* (Cambridge, Mass.: MIT press, 1997).
9. Owen Flanagan, *The Science of the Mind* (Cambridge, Mass.: MIT Press, 1991), p. 228
10. *Ibid.*, p. 241.
11. Peter Strawson, *Individuals* (London: Methuen, 1977), p. 103.
12. In cultures outside the West, "personality" as understood from a linguistic point of view does not exist, or exists in such a radically different way that it is senseless to claim any meaningful comparison. See the discussion in Gerald M. Erchak, *The Anthropology of Self and Behavior* (New Brunswick: Rutgers University Press, 1992), pp. 8ff. Erchak notes in social sciences overlapping and competing definitions of "personality" and "self", and a shift in the use of "personality", Generally, it now comes to refer to "a more 'inner' theoretical concept, inaccessible to direct observation, whereas the 'self' is conceptualized as something 'presented' to the community at large and thus accessible ... through behavior observations, autobiographical accounts and so on," and that psychocultural studies is now replacing it with "self" or talk of a egocentric and sociocentric selves. pp. 9, 11. For the account of person given here see also my "Identity, Personhood and Religion in Caribbean Context", in *Nation Dance*, ed. Patrick Taylor (Bloomington: Indiana University Press, 2001), pp. 138-152.

13. *Khasya* means to know, and *sam* means exact. Interestingly, *sankhya* also means number since this school counts the number of categories (25 in all). It is nontheistic in its narrow philosophical reference, and yet its founder Kapila is a theist. The categories it mentions are found also in the Vedanta schools, thus leaving the impression that there is theistic Sankhya as well. For more on this, see for example, the discussion by P. T. Raju, *Structural Depths of Indian Thought* (Albany: SUNY press, 1985), p. 304f.
14. According to his friend Maurice Drury, Wittgenstein considered Kierkegaard to be the greatest 19th century philosopher, and in a conversation recounted by Waismann, he linked his own ideas to Heidegger and Kierkegaard. See M.O'C. Drury, 'A Symposium:..' in *Ludwig Wittgenstein: The Man And His Philosophy*, ed. K. T. Fann (New York: Delta Books, 1967), p. 70, and is mentioned also by Allan Janick and Stephen Toulmin, *Wittgenstein's Vienna* (New York: Touchstone Books, 1973), pp. 25f. For the conversation see F. Waismann, *Ludwig Wittgenstein and the Vienna Circle*, ed. B. F. McGuinness (Oxford: Blackwell, 1979), p. 68, which is cited by Cyril Barrett, *Wittgenstein and Religious Belief* (Oxford: Blackwell, 1991), p. 22 who identifies it as a scandal to British Wittgensteinians. In a letter to his friend Paul Englemann written in 1921, Wittgenstein recommended Tagore's *King of the Dark Chamber*. Englemann noted that Wittgenstein gave a copy to his sister, that the book was listed as one of his favourites, and that in 1927 and 1928 meetings of the Vienna Circle would read out poetry from Tagore rather than discuss philosophy. See Paul Englemann, *Letters from Wittgenstein* (Oxford: Blackwell, 1967), pp. 44-47. Interestingly, the influence of Tagore on Wittgenstein remains unexplored, notwithstanding the publication of the letter (1967), mention of it by Janick and Toulmin (1973), and discussion by Ray Monk of Wittgenstein's translation of Act II of the play in Ray Monk, *Ludwig Wittgenstein: The Duty of Genius* (New York: Free Press, 1990), pp. 408-410.
15. Sören Kierkegaard, *Works of Love* (New York: Harper Torchbooks, 1962), p 199.
16. Rabindranath Tagore, *Sadhana* (Tucson: Omen Communications, 1972), pp. 71f.
17. Englemann, pp. 143f.
18. The point about a richly textured life and the exercising of capacities is one that Paul Holmer at Yale (my teacher) emphasizes. See Paul Holmer, "Wittgenstein and the Self", in *Essays on Kierkegaard & Wittgenstein*, ed. Richard H. Bell and Ronald E. Hustwitt (Wooster: College of Wooster, 1978), pp. 27f.
19. Barrett, pp. 30, 17f.
20. Ibid., pp. 20, 25.
21. Ibid., p. 21.
22. Ludwig Wittgenstein, "A Lecture on Ethics", in *Philosophical Occasions 1912-15*, ed. James C. Klagge and Alfred Nordmann (Cambridge: Hackett Publishing, 1993), pp. 43f.
23. Waismann, p. 68, also Barrett, p. 22.
24. Rabindranath Tagore, *The Religion of Man* (London: Unwin, 1931), p 40.
25. *Ibid.*, pp. 41, 33, 91.

26. Rabindranath Tagore, *King of the Dark Chamber* (Madras: Macmillan, 1980), Pocket edn. p. 199.
27. *Religion of Man*, p. 83.
28. *Ibid.*, p. 154.
29. Rabindranath Tagore, *Lectures and Addresses*, ed. Anthony X. Soraes (Delhi: Macmillan, 1995), Pocket edn. p. 93.
30. *Ibid.*, p. 87.
31. *Religion of Man*, p. 11.
32. Ibid., pp. 82, 70f, 85.
33. *Lectures and Addresses*, pp. 84f.
34. *Ibid.*, p. 73.
35. "The Last Testament of Bertrand Russell", *The Independent*,(London), 2,215 (November 1993), p. 24, and excerpted by Peter Caws, "Identity: Cultural, Transcultural, and Multicultural", in *Multiculturalism: A Critical Reader*, ed. David Theo Goldberg (Oxford; Blackwell, 1994), p. 385.
36. *Lectures and Addresses*, p. 74.

EXPLORING MIND AND CONSCIOUSNESS

EMPIRICAL APPROACHES

An Experimental Approach to Conscious Intelligence

G.C. RAY

I. Introduction

Consciousness is as old as our civilization though scientific investigation about its origin and nature has been recently attempted. There has been considerable emphasis on body-consciousness or the consciousness of human being because of its importance in anaesthesia and surgery as well as the effect of different drugs. The clinical research has its own importance though in this paper we will mainly concentrate on the nature and genesis of consciousness from the point of view of *ultimate reality* and evolution.

Recently Chalmers (1996) has divided the problems on consciousness into two parts: *easy* and *hard*. The ability to react, discriminate, categorize, integration of information by the cognitive system, control of behavior *etc* are, in his opinion, *easy* problems. Though these are certainly not easy, but one may think of the scientific planning of experiments which are possibly within our limitations. The real *hard problem*, he emphasizes, is the *generation* of *experience* or *qualia* or simply the subjective fold of human beings (or of any other species, taking into account all his integral domain of memory, intelligence, subjective manipulation of images, analysis, decision-making etc). Many researchers in the area of artificial intelligence (AI) suggest that the subjective fold is generated autom-atically when the complexity of the system is high enough, whereas many are skeptics about this view. Chalmers himself could not find any alternative but dualism and suggests *information* as another funda-mental entity apart from matter and energy. It is interesting to note that the same problem was faced in the philosophy of the *Upaniṣads* where it was found that matter and energy, listed as five elementary entities, may be correlated to the universe but conscious-ness remains an unique problem. It was therefore considered one of the *indirect* attributes of ultimate reality, the *Brahman*. Swami Prabhavananda (Prabhavananda and Isherwood 1994) writes of the cosmology of the *Gita*,

'*Upanishads* say that Brahman is Existence, Knowledge and Bliss; but these are not attributes. Brahman cannot be said to exist. Brahman is Existence itself. Brahman is not wise or happy, but absolute Knowledge, absolute Joy.' It may be noted that nothing is mentioned here about consciousness. One therefore may ask 'is Brahman consciousness?' The answer is 'yes' if consciousness is taken as the pre-requisite for any knowledge, past present future, or an integral form of these three. Indeed unconscious knowledge is the attribute of an automaton. We define *chit* as conscious intelligence and in this paper *chit* is used in this sense.

In another view, body-consciousness or simply consciousness is, possibly a tool developed in the process of evolution to assist survival. It should not have more importance than the development of a long neck, slender leg, club-like tail or visual organ. Body-consciousness sustains *experience* or the *qualia* of Chalmers. It gives an m-property or mental property to every p-property or physical property; for example, the electromagnetic radiation of wavelength 5890 angstrom-unit is a p-property and 'yellowness' is the m-property. There is a one-is-to-one correlation between the two, but otherwise these are unbridgeable.

M-properties also possibly followed a pattern in evolution which is as yet inadequately studied. A prey animal will like to see the tiger (predator) in bright yellow with black stripes on it on a background of green for maximum visibility. This enhances its survival potential. Other species, for whom the tiger is not a predator, may not need the same color-contrast, and they may see it in different shades. That is, the color one sees does not give the true shade of the world. It may be interesting to study from the point of view of evolution why blood appears red, soil grey and grass green.

Body-consciousness may thus be deceived easily. Imagine a huge building, in an ever-dark forest where light is not known, with an infinite number of rooms each enclosing one species. Suddenly one day, a spark ignites the deposited sulphur on the floor of a room and dim yellowish light appears (like consciousness). The inhabitants, already possessing visual organs, will find in surprise that they can do all their work much better and with improved clarity. They will also see walls, floors, furniture in yellow and will visualize the universe as yellow which is not necessarily true. Body-consciousness forces one to see the world in a particular way enabling calculation and analysis for one's survival, but may not give the true picture of the universe.

Possibly consciousness has been there from the very beginning of evolution in a dim form. If a huge aquatic animal, driven by a predator, can have a faint mental picture of the last attack like that in delirium, it enhances its survival potential considerably.

Though survival is the pivot of any evolutionary process, it is true that the latter has other dimensions as well. While the survival aspect of consciousness may distort the world-picture to enhance survival; it is likely that the non-survival aspects (mainly the domain of aesthetics) will be relatively free from this feature (discussed in detail afterwards).

Consciousness, as one understands it in *Homo sapiens*, is an intense form compared to what it was in that aquatic animal, described above. This higher intensity has both advantages and disadvantages. The advantages are in better calculation, analysis and decision-making; the disadvantage is that the animal cannot easily go to the restful state (sleep) because some physiological subsystems may remain excited. It is, therefore, required to 'switch-on' or 'switch-off' consciousness whenever necessary. It is likely that this 'on-off' mechanism appeared at the latter stage of evolution and at present an intense scientific effort is being made to identify this 'seat of consciousness'.

Are *chit* and consciousness, thus defined, the same or different entities? Is it possible to know *chit* through body-consciousness? Statements on this question are largely disappointing. The *Rigveda* says, 'It is not by our human nature that we know of Him.' Some more statements by Sankaracharya are as follows: 'So long as the idea of the Self is identified with body-consciousness, *i.e.*, till the realization of Atmana, the world perception, the reality of the world is valid,' 'Brahman, although not knowable (by the intellect) is yet not unrealizable,' 'Brahman shining in the depth of heart cannot be realised by the external senses,' 'The Self is without adjuncts, indescribable, without parts, without qualities and pure which the mind and speech do not reach.' In another place he says, 'The Self is conscious even in deep sleep as well as in waking,' if the Self here is replaced by one of its attributes *chit*, then *chit* exists even in deep sleep where, as we know, consciousness does not, clearly suggesting that the two are different. The barrier between the two is ignorance or *avidya* and once that is removed, 'the Self reveals itself like the Sun when the cloud is removed' (Sankaracharya). This type of description has created the impression that the barrier in between is like a mountain and climbing it is extremely difficult; once the

climber is at the top of the mountain, a completely different landscape comes into view where every inch of detail differs from all our previous perceptions.

Sankaracharya mentions that this path of climbing is the path of purification, but Ramakrishna clearly states that 'pure mind, pure intellect, pure soul — same thing' (*Sri Ramkrishna Kathamrita khanda* 4, *adhyay* 19, *parichhed* 2). This possibly suggests that *chit* is the purer form of body-consciousness. Ramakrishna further clarifies the path of attaining purity, 'form, sensibility, smell, touch . . . nothing will remain in the mind – then only it will be pure' (ibid. 4/7/5). Obviously the mountain remains equally high with the consolation that it is the body-consciousness which is transcended to *chit*. Two different entities would have been difficult to handle from the point of view of evolution.

In order to have experimental results on the nature of *chit*, we required subjects who have climbed up to a considerable height of that mountain. Such persons, though not rare (as I could sense from my talk with a few monks), are usually reluctant to enter into a large Faraday cage within the alien environment of a laboratory where sight, sound and smell are not conducive to their familiar mode of concentration. In spite of our best efforts, we could not collect subjects and hence we turned towards non-survival aspects of consciousness. Frequently it is seen that village boys are fascinated with certain tunes or lyrics and they pursue them disregarding the survival aspects of the hard realities of life. They are basically unattached like Tarapada of the famous short story *Atithi* by Tagore. The innocent girl (Kurani) of *Malyadan* may also be mentioned, who, for her serene love took the journey to her beloved absolutely forgetting the thirst, hunger and plague on the way. May the drive of non-survival aspects of consciousness dominate over survival instincts? The concept of evolution may not agree with this, but such examples are abundant in every culture. In a sense, Sri Ramakrishna is also an example. When he was only six, he fell unconscious (interpreted now as his first *samadhi*-like state) in a cornfield seeing a row of white cranes on the border of black clouds, a familiar scene in rural Bengal. One of his biographers, Romain Rolland, comments that his journey towards ultimate reality possibly came through aesthetic pathways (Romain Rolland 1965). We concentrated on aesthetic appeal with appropriate subjects for our search on the non-survival aspects of consciousness and possibly on *chit*. The additional advantage here was that the experimental results could be repeated as many times up to satisfaction, a rigid demand of scientific methodology.

Most of the subjects of these experiments were honest, upright and disciplined, they also used to maintain a low profile and were largely introverts. Because of these selection criteria, good results were obtained from more than 50% of them, out of a total of about two dozen participating in the experimental sessions spanning more than five years.

The methodology and equipment used are explained in section II. Details of preparation of the subjects and the planning of experiments are also given here. The experimental results are placed in sections IV, V and VI. The section IV deals with the signature on electroencephalogram (EEG) when consciousness becomes 'cloudy' for any reason – aesthetic appeal, childish amusement or meditation. As is seen during *grand mal/petit mal* epilepsy, here also the typical signature is an isolated spike preceded and followed by isoelectric lines or low-frequency low-magnitude waves; indeed EEG cannot differentiate the clouding of consciousness by different methods and the singular representation is the spike or sometimes trains of a few spikes not necessarily at equal intervals. This fact has been known for more than half a century. An attempt is made in this section to provide a theoretical framework for this phenomenon. In section V, the change in the complexity of a system, here the central nervous system (CNS), is discussed in terms of a single parameter and an experimental procedure to capture it using the chaos model. The output of the model is used to correlate the physiological symptoms of Sri Ramakrishna during *samadhi*. In section VI, heart rate variability (HRV) model is used on ECG to know more about the autonomic system when consciousness approaches towards ultimate reality. In the last section, section VII, a probable genesis of consciousness is discussed taking into account the experimental results and all other contemporary views.

We have provided here a completely non-mathematical treatment of the subject. It is attempted to keep the essence of ideas but more serious readers must go through the references for the complete details of mathematical models. Readers not willing to take that trouble, must read section III carefully where scientific/engineering terms like signals, periodicity, spectrum, filters, impulse etc are explained in non-mathematical language. Some other more complex terms like poles and zeroes, auditory spectrum etc are explained in appropriate places.

II. EXPERIMENTAL ARRANGEMENT

EEGs were picked up in a room with all lights off where the subject, child or adult, had his eyes closed. For all the experiments, subjects were seated in a reclining chair which could be adjusted to their comfort. The whole arrangement of seating and data collection, using the head-box of EEG, was placed inside a large Faraday cage. It had acoustic as well as electromagnetic insulation with revolving doors in the side and roof. The EEG machine was placed outside the cage and the signals were carried through a cable, passing through the walls, to the adjacent room. EEGs were initially stored in analog form in TEAC HR-10E and TEAC MR-10 (4 channels each) instrumentation recorders. The data were observed in a signal analyzer AD-3521C which could digitize the data, 15 bits per sample at 400 Hz. All the 8 channels, each of 45 minutes duration, were analyzed carefully and the significant portions marked for further analysis. These portions were digitized in a PC using 12-bit analog-to-digital converter. These files were used for the various models. All the experiments were carried out at midnight to avail of the advantages of solitude.

Preparation for the experimental session was very important. Subjects were asked to choose their beloved music, usually old favourites worked best. The subjects did not hear that music for a few months so that the appeal on hearing it during the experiment should be strengthened. On the day of the experiment they were asked to remain introverted and not to talk with anybody from mid-day up to the session at midnight.

Initially experiments were carried out on music lovers known for their knowledge and appreciation of Indian classical music. Though the signature on EEG, the spikes, were obtained the numbers were less and the spikes were of smaller magnitudes. As we proceeded, we found that devotional music with devotees as subjects gave the best result; possibly the aesthetic appeal along with the stress of devotion helped the subject to reach the epoch moments. Sometimes some subjects were advised to use imagery; for example, 'a singing mendicant passing through a vast meadow and merging with the horizon in evening twilight' is very helpful for the subject to transcend along with aesthetic appeal.

Apart from aesthetic appeal, we also used highly amused states of children which we assumed equally pure and serene. An eleven-year-old boy, certified by the school as very imaginative, was one such subject. The highly amused state was created by playing back a cassette containing his baby-talk recorded earlier by his parents during his childhood in an animated family episode. The child used to burst into uncontrollable laughter on hearing his own baby-talk and the muscle spasms were excessive. We had to play the cassette many times till the child could control his myoclonus. The elder sister of this subject, 16 years old and who also had a part in that family episode, took part in another session. Another child, 6 years, was used with nursery rhymes recited by him earlier. Among the meditators, one had 20 years' practice in a certain form of meditation, carried out in his family for the last three generations. Among the music-lovers, several subjects used western classical (mainly our students) including Beethoven's symphonies. The old principal of our campus school, a Kashmiri brahmin (he is no more) and devoted music lover, used folk-song-based *Kabir bhajan* recorded by him fifteen years ago from a travelling musical troupe; he valued this cassette even to his last days.

III. A FEW DEFINITIONS

If you talk in front of a microphone and connect the output of the microphone to an oscilloscope, you may see a time-varying plot of a line changing very quickly on the screen. This is what is called a signal, sine wave or square wave being simple examples. For a vowel sound, try any sustained vowel, you will see not the sine wave but a complex wave. If the oscilloscope is a storage type, press the storage button and the image will be fixed on the screen. Look carefully, the wave shape may be complex, but it is repeating the shape after some time-intervals called pitch period. For normal male human beings, this pitch may be anywhere between 3 to 8 milliseconds. When earth vibrates, water flows, wind blows or any change occurs anywhere, a time-varying electrical signal may be obtained using devices called transducers.

The concept of spectrum is very important, particularly here to understand the models. If you mix three sine waves, each of different magnitudes, you will get a complex signal but again periodic. Looking at this signal, it is very difficult to identify the magnitudes and frequencies of the three sine waves. On the other hand, if you record the howling of the wind, the murmuring of bamboo in silent summer-noon or the chirping of a sparrow, every time you get a signal, but

you do not know the frequencies comprising it. Why it is necessary to know? If you know the frequency and its variation in the chirping of the sparrow, you may know more about the source of the sound, the variation in the tension of the tissue vibrating somewhere inside the tiny mouth-cavity. In the case of the human voice, several cavities in addition to the vocal cords and vocal tract, are taking part. Each has its own natural frequency of vibration. These parameters are very individual to each and every human being, *i.e.*, from the frequency-patterns you may even identify the person and this is called speaker identification in signal processing.

A mathematician, Jean-Baptiste Joseph Fourier (1768-1830), first provided the mathematical framework to find out the spectrum from a signal. If it is used in that above-mentioned signal of the mixture of three sine waves, you may clearly see three peaks in the spectrum (here the x-axis is frequency) suggesting that three sine waves are mixed; not only that, the magnitudes of the sine waves are proportional to the height of three peaks. Laplace transform, named after Pierre Simon, Marquis De Laplace (1749-1827), is more involved and takes DC transiencies into account. The digital version of Laplace transform is called z-transform. These three still now make the backbone of signal processing.

Most of the signals may be expressed in mathematical form as a function of time. If the mathematical expression is very complex, in general the signal will also look complex, but the signal obeys the laws of the time domain. There is only one type of signal which does not obey any law of the time domain; its magnitude at any instant of time is independent of any number of previous magnitudes in the past instants. This is called a random signal and it plays an important role in the present work. Using the Fourier expressions, one may find out the sine waves making the random signal. The result is particularly interesting; it shows that an *infinite number of sine waves of all frequencies are mixed up to make the random signal, i.e.*, the spectrum is a horizontal line at certain magnitude along the frequency axis. But this nature of the random signal is true if the length of the signal is large enough. If the length of the random signal is small, the Fourier spectrum does not show equal magnitude for all sine waves; as the length of the signal increases, the spectrum becomes more and more of constant magnitude, a concept used in the next section.

Another interesting concept is impulse. It has infinite magnitude, zero width but unit area under it. One may say that impulse also

does not obey any law of the time domain, because infinite height and zero width are all theoretical impossibilities and one may wonder how the area under it may have a distinct value of unity. But this fascinates mathematicians and puzzles us. It is more interesting to note that its spectrum is also like that of the random signal, i.e., of constant magnitude along the frequency axis. Now think of an infinite number of sine waves of the same magnitude; if you mix these, the result is an impulse!! You may be amused but it is conceptually true – think more on it, because we have used the concept in this work.

Both random signal and impulse are termed in this paper as *transcendent* signal, because none of these obey any law of causality of time domain.

In practical cases, impulses are generated by electronic gadgets. Since infinite height and zero width are both impossible, a sufficient height with minimum possible width is granted as practical impulse. Infinite height also indicates infinite rate of rise, this is also impossible because of limited slew-rates of all electronic chips or limited speed of biochemical reactions going on inside the cerebral cortex. The spectrum of the practical impulse will also show the summation of a large number of sine waves but not infinite. In many cases, this large number of sine waves may not be necessary for a particular purpose. Hence the signal may be passed through filters. A lowpass filter will pass out only low frequencies up to a certain cut-off, whereas highpass filters pass out high frequencies only above the cut-off.

IV. WHY AN IMPULSE IN EEG MEANS TRANSCENDENCE

Impulses in EEG due to aesthetic appeal were first noted in 1937 by Critchley (1937). Since then the topic has been given the technical name musicogenic epilepsy. Work on it was mainly carried out during the sixties and seventies of the last century. Servit *et al.* (1962) reviewed twenty cases where seizures were triggered by music; many of the reports indicated that the induced seizures were located in the temporal lobe and probably arose from foci in the auditory cortex. That musicogenic epilepsy may be related with something 'devotional' is also mentioned earlier. Poskanzer *et al.* (1962) reported that only a discrete frequency band of *church bells* caused musicogenic epilepsy and Marsden *et al.* (Laidlaw *et al.* 1982)

mentioned a patient whose *grand mal* seizures were initially provoked by certain types of classical orchestral and church music, and who could be induced to have a fit by Beethoven's Fifth Symphony.

As an explanation of this phenomenon, a theory is given as follows: when the consciousness transcends the time domain, it is no longer bound to any causality of the time-frame; following psychophysiological parallelism, the corresponding signal (the p-property) will also show transcendence (Ray and Kaplan 1994; Ray 1995). We have already seen that there are two types of transcendent signal – the random one and impulse, with the limitation that the duration of the random signal must be very long to exhibit the property of transcendence. But in the experiments of musicogenic epilepsy or in our experiments, the consciousness transcends the time domain with quick re-entry and in many cases (as in *petit mal* epilepsy of children) it does not seriously disturb the continuity of consciousness just as blinking of the eye does not disturb the continuity of vision. Hence the appropriate representation will be an impulse and not a random signal. In EEGs, the durations of spikes (impulses) ranging from 43 to 86 milliseconds are considered characteristic. These durations are likely due to the neuronal network of the cortex and its biochemical response-time.

Fig.1 shows some of the EEGs containing trancendences picked up from some subjects. (a) and (b) are from a subject under the spell of music; this type of signal started appearing after about half an hour (duration of each session was 45 minutes). Signal (d) was from the child in a highly amused state and (f) and (g) were due to meditation; spikes in the latter cases are usually smaller and fewer in numbers. However genuine spikes are identified not by their height but by their isolation (i.e., they should be preceded and followed by isoelectric lines or low-magnitude slow waves) and sharpness (the thumb rule says, if it pricks it is genuine).

The music was also recorded in one channel. After preliminary analysis, the portions of the EEGs containing numbers of spikes were termed *epoch* periods (usually of the duration of few seconds), the rest of near-normal EEGs were *non-epochs*. In most of the cases, the music and data were played back to the subjects and they identified that epoch periods coincided with the moments they found extremely pleasant or they were 'swept away'.

At this stage we wanted to investigate in detail what are the factors responsible and driving the subjects to be 'swept away'. It was

anticipated that the recording of music in the cortex might provide some clue. Signals, whether of music or anything else, are not recorded directly because it takes a large amount of memory — instead salient aspects are extracted and recorded; this is true for digital electronics and must be so for the cortex. One of the simplest methods is to find out the dominant frequencies by Fourier spectrum and store these values. Might this be done within the cortex also? Researchers had serious doubt. From the preliminary study of the hearing organs of *Homo sapiens* and other mammals, it was known that these also extract dominant frequencies but in a different way. These values are sent to the cortex via the auditory nerves and these sets of frequencies also interact with the neurons possibly in a convolutional manner. The whole process has been developed by the process of evolution and is likely to be different from the Fourier algorithm. In the eighties and nineties of the last century, considerable attention was paid to this and the result was a different spectrum than that of Fourier, the *auditory spectrum*. We concentrated on this auditory spectrum and its interaction with the *response function* of the neurons to know why a particular portion of a piece of music appeals and in general the reason for transcendence.

Sound waves vibrate the basilar membrane of the cochlea producing standing waves whose amplitudes peak at specific frequencies; the distance along the cochlea has a logarithmic dependence on the frequency of the input stimulus. The mechanical movement is transduced to an electrical signal by an ionic current flow into the hair cells, which are controlled by the bending of cilia caused by the motion of the basilar membrane. The electrical potentials generated across the hair cell membranes are finally conveyed by auditory nerve fibers in the form of neural spikes to the cochlear nucleus. The auditory nerve response is usually modelled as a Linear Inhibitory Network or LIN. All these stages (some smaller ones have been dropped for clarity of explanation) may be represented by blocks and the input-output relationship (in the frequency domain the input-output relationship is called the transfer function) of each block was determined experimentally on animals (ferrets and cats). These transfer functions were concatenated to get the auditory spectrum. We followed the model by Wang *et al.* (Wang and Shamma 1994, 1995) which takes into account the cochlear and differential cochlear filter in a rather realistic manner.

Fig.2 shows the auditory spectrum of a sustained vowel. If it is compared with the Fourier spectrum of the same sound, it will be seen that the peaks are occurring at different frequencies. It is, therefore, the way evolution chose to extract information from sound signals. Now this signal is to interact with the neurons.

In the primary auditory cortex (A1), neurons exhibit certain organizational characteristics. Each nerve cell tends to be tuned to a specific range of frequencies (Merzenich *et al*, 1975) creating the tonotopic axis which has a 2-D nature with many cells tuned to the same frequency lined up along the isofrequency planes that run perpendicularly (see the detail in Shamma *et al*, 1993). The response function of the neuron is considered (with some experimental support) a variation of the negative second derivative of Gaussian and the interaction with auditory spectrum is inner multiplication or convolution.

Fig.3 shows the representation of two musical signals on the cortex. The x-axis is the tonotopic one with log-frequency, y-axis is the magnitude and z-axis is the phase which is due to the particular distribution of inhibitory and excitatory neurons in A1. Note the difference in these two cortical representations. The first one has a sine wave-like distribution (it is called *ripple*) of magnitude, whereas the second one does not have it. Indeed the first one represents epoch period and the second one, non-epoch (Das and Ray 1998).

In order to explain it in a better way, we show in Fig.4 two more cortical representations, the upper one is of a monotone (sine wave at 500 Hz) and the lower one is of noise. None of these two sounds are pleasing and hence the distribution of magnitudes are far from ripples.

Will this always be the case with something highly pleasing and not-so-pleasing? We attempted to find out whether music of high aesthetic quality always gave rise to sine-like (SL) functions and non-epochs to non-sine-like (NSL). Some *ragas* from Indian classical music were used. According to the structure of *ragas*, each uses only certain notes (not all) in a definite pattern of ascent and descent called *aroha-abroha*, and evokes highly pleasant specific sentiments. Many of these musical phrases have been refined through the centuries and passed from generation to generation. These are mentioned here as *characteristic-raga* (ch-raga). Maestros start with these then go on to imaginative elaboration (*vistara*), but repeatedly come back to ch-raga. In this experiments, five *ragas* (all instru-

mental) were selected. From each 36 frames (18 in ch-raga and 18 in vistara) were digitized, having 8192 data points in each. Neural representations were found out for each and the result is shown in Table-1.

In each *raga* the best SL and worst NSL were chosen and distance between the two was calculated. In binary, this distance was divided by two. If a particular frame has the distance less than this, it was considered in SL group, otherwise it belongs to NSL group. In ternary, the distance between the best SL and worst NSL was divided by three and three groups, SL NSL and mixed (M), were created. It was expected that the ch-raga, which is refined through centuries, will have maximum number of SL and *vistara*, maximum NSL. It may be verified in the table that indeed this is the general trend, except two cases (*Tilak-Kamod* and *Yaman-Manj*) where subjects identified *vistara* more pleasing aesthetically.

This result has a profound significance. The sine wave on the linear frequency axis is transformed (by Fourier) as a single large peak on the time-axis, something looking like an impulse. But *ripple* is a sine wave on the log-frequency axis; on transform, it becomes a slightly broad-based impulse as shown in Fig.5. That is, deeper aesthetic appeal induces on time-plane that transcendence again, which is the gateway or escape-route from all causalities of the time-domain.

V. CHANGE IN COMPLEXITY OF CNS DURING TRANCENDENCE

In this chapter, we examine the change in complexity in central nervous system (CNS) following the chaos model. This complexity, from the point of view of signal analysis, may be comprehended in different ways. Assume that several machines, mounted on the floor of a building, are running as seen in workshops. Each machine has different speed of rotation, different torque and different types of mutual coupling. The entire building is a complex system and nobody is allowed to enter into it, but the complexity is to be measured. One may mount a strain gauge on the floor just outside the door and may get a signal proportional to the vibration of the floor following the methods of an earthquake engineer. It will be noted that the more complex is the vibrational pattern, the more complex the signal will be. The method used here is almost the

same; no details are known of how the CNS is working particularly during transcendence, only EEGs may be picked up without disturbing the subject. There are again different ways to interpret the complexity of the signal producing definitions of correlation dimension (*cd*) or fractal dimension. We have used here *cd* which may be explained as follows.

Assume that the length of a riverbank is to be measured for some constructional purpose. The total length, say, is near 500 meters. Since it is that much only, one may use a scale of one meter and finish that insignificant task. But say the length of the scale is reduced to one centimeter and the instrument-room does not find a longer scale. This time the scale will go along the curved edges of each boulder and pebble and the total length may be several times the previous measurement; *i.e.*, the smaller is the scale, the more will be the total measured length. If the length of the scale is reduced to one micron only, the total measured length will be considerably higher because the tiny scale now will go round even sand, dust and similar particles. Theoretically the scale may even be shorter, such that it may go round molecules and atoms. Since subatomic particles have both particle and wave properties, we really do not know how they 'look' except through their conceptual models. One may easily see that the simple job of measuring the length becomes a philosophical impossibility. But the practical engineer will not go up to that. If he is very much a lover of truth, he will produce a graph - the x-axis indicating length of scale and the y-axis, the total measured length. When the scale is of one meter, the measured length is 500 meter and let us join these two points by a straight line. Next the scale is of one cm and the measured length, say, is 1500 meter and these two points are also joined by another straight line. It is easy to visualize that the slope of the second line will be higher. This is in general the rule. The more complex the line, the higher will be the slope.

The correlation dimension (*cd*) is measured in a similar way on the EEG signal. The complete mathematical formulation is fairly complex and will be available in Parker and Chua (1987), Farmer, Ott and Yorke (1983) and Mayer-Kress (1986). A typical graph of more and more slope in M-dimensional phase space is shown in Fig.6 and the maximum slope represents *cd*. Here *r* of x-axis is a parameter like that of the scale of the previous example, and is expressed in log, and the y-axis, again in log, represents the total

count.

The result of this analysis is given in Table 2. For four subjects, the variation of cd is given in the first two columns. Initially sections of signals, each of about 20 seconds (4096 data points), were marked as epoch and non-epoch periods. For the subject 1, the average *cd* was 3.27 in nonepoch, but it was reduced to 2.48 in epoch. The variations in the other subjects followed similar pattern.

This complexity of the EEGs was also measured by another technique called r-parameter (Ray, 1995). The idea followed a paper published by May (1976). This author used mathematical expressions where a single parameter could be changed to make the signal more or less complex. This signal was basically compared with EEGs to arrive at the value of r-parameter. The result is shown in Table 2 in third and fourth column. Here also it is seen that r-parameter is reducing during transcendence.

The significance of lesser complexity of the CNS during transcendence was initially unclear. It was not known whether it is responsible for the calmness associated with meditation. From the point of view of signal processing, the lesser complexity may be the result of increased crosscorrelation between the EEG signals coming out from different parts of CNS. Whether this increased coherence between different parts of CNS is really taking place or not was the next point of our study. The result is shown in Table 3. For the subject-1, the cross correlation between two non-epoch segments of EEG was 0.277, whereas during transcendence it was reduced to 0.007 when the epoch segment was compared with another non-epoch segment of another channel. This trend was seen for all other subjects in all other channels taken for study. In short the correlation between different parts of the CNS is rather poorer during transcendence. Sometimes it is claimed by some spiritual groups that the coherence between different parts of the brain and body is higher during its journey towards the ultimate goal – an idea we found not supported by experimental results.

What then may be the reason for the reduced complexity of the CNS during transcendence? We studied in detail the symptoms observed in Sri Ramakrishna during, before, and after, his *samadhi*. Some are mentioned here: (i) *high body temperature*: he used to have this during the earlier stages of his practice, and would remain immersed in the water of the river Ganga up to the neck, with a wet napkin on his head, to combat the temperature (*Kathamrita* 3/14/

2), (ii) *pain in the belly*: during his vision in a mentally heightened state, this pain used to come, usually the next day. Sometime the pain used to be so acute that Ramakrishna used to avoid further visions (Kathamrita 4/23/9). (iii) *electromagnetic radiation from the body* – the chest used to be reddish (*Kathamrita* 4/24/3) and a sort of halo used to be seen by the naked eye by others (*Kathamrita* 1/7/3), (iv) *spatio-temporal sense vanishes:* in one scene just on the verge of *samadhi* he says, 'I am looking at you, but when you have come, where you have come ? can't remember.' (*Kathamrita* 3/14/1), (v) *locomotion is lost:* in another situation, just after coming out from *samadhi*, he is not able to walk, the power of locomotion is lost (Kathamrita 3/4/3), (vi) *tremor of thumbs and body:* during the onset of *samadhi* these were also observed (*Kathamrita* 4/19/3).

Observers also found that, during *samadhi*, the power of deglutition is lost; milk rolled out if poured in the mouth. The loss of locomotion is also mentioned. Possibly the subsystems of the human body gradually switch off during the journey towards the ultimate reality; and this may reduce the complexity of the CNS. In the previous example of the work-floor if several machines are switched off, it will reduce the complexity of the signal picked up by the strain gauge.

The process of dropping the subsystems, from gross to subtle during the journey towards ultimate reality, is the core of the philosophy of *Upanishads*. It is the path of *neti* ... *neti*, not this ... not this and leave all, one after the other.

VI. EVOLUTION AND 'LANDING ON ULTIMATE REALITY'

In this last section of experimental results, we studied the autonomic systems of the subjects during their transcendent states. The model used is called heart rate variability or simply HRV. Swami Vivekananda once wrote in *Rajayoga*, that many systems of the human body may be controlled through respiration. Indeed many subsystems of the body impose their periodicities on the periodicity of respiration. In the ECG, the large spike-like (QRS-complex) magnitude has at its peak the point R and the interval from one R to the other is called R-R interval. If you look at the ECG closely, you will find that the R-R intervals are all unequal. These are reducing gradually during inspiration and increasing during expiration.

Hence respiration is imposing its periodicity on the hear rate. Blood pressures also vary in the human body around a set-point; if the pressure is higher than the set-point, mechanisms in the body pull it down, if the pressure goes down below the set-point, another set of mechanisms pull it up. This process has the time constant of about 25 ms and it also imposes its periodicity on the R-R intervals. Similarly, thermoregulation (the mechanism to keep the body-temperature constant at set-point) also imposes its periodicity. If the spectrum of R-R intervals, taken for several minutes, are plotted, it will look as shown in Fig.7.

Here the x-axis represents frequency and the y-axis, power in different peaks. The first peak is very near to the origin and at present we will not use it. The second peak represents the power in the variability of blood-pressure control and the third peak, is called respiratory sinus arrhythmia or RSA. The frequency of this peak is also the frequency of respiration. This peak is also considered a 'marker' of vagal tone. Hence HRV spectra give a composite picture of the internal system of the body involving respiration, blood pressure, body-temperature and some other aspects. These also constitute the autonomic system of the human body.

The autonomic system has two main mutually opposing subsystems: sympathetic and parasympathetic. Some of the symptoms of sympathetic stimulation are: (i) increased cardiac rate and coronary flow, (ii) increased respiration rate (tachypnaea), (iii) inhibition of intestinal peristalsis and holding-up of the contents of the gastro-intestinal tract (GIT), (iv) cutaneous vasoconstriction and increased body temperature, (v) increased blood sugar level (hyperglycaemia), (vi) decomposition of fat (lipolysis) and (vii) piloerection (hair-rise).

The effects of parasympathetic stimulation are the opposite: (i) decreased cardiac rate and reduced cardiac output, (ii) reduced respiration rate, (iii) increased secretory activity of the glands of saliva, stomach and pancreas and (iv) increased motility of the gut resulting in increased movement.

Function of the autonomic system is reflected in the HRV spectra. For example, higher power in the second peak and higher ratio between second to third peaks indicates a sympathetic burst in the autonomic system. For the normal body and mind, a perfect autonomic balance is required; its unbalancing also indicates diseases and since the early seventies HRV spectra has been used for the

assessment of autonomic functions as well as for various pathologies like hypertension, diabetes and myocardial ischaemia (Hyndman *et al.* 1971; Kitney and Rompelman 1980; Akselrod *et al.* 1981; Billman and Dujardin 1990; Bianchi *et al.* 1990).

Some of the simultaneous EEGs and ECGs, picked up from our subjects, are shown in Fig.8. Here (a) and (b) are the normal EEG and ECG during non-epoch periods. The intervals (c) & (d) and (e) & (f) represent epoch periods with lots of spikes on the EEG and some morphological change in ECG which are discussed elsewhere (see Ray, Kaplan and Jovanov 1997, 1999).

R-R intervals from ECGs of both epoch and non-epoch periods were calculated and spectra were obtained following a procedure given by Mainardi *et al.* (1995). The typical set of spectra for epoch (full line) and non-epoch (dashed line) periods are shown in Fig.7. It may be noted that the power in the second peak is more during epoch period, also the ratio between the second and third peaks is larger here; this clearly indicates a sympathetic burst in the autonomic system.

Some of these symptoms (of sympathetic burst) like increased body-temperature, pain in the belly (may be due to inhibition of intestinal peristalsis and holding-up of the content in the GIT) were indeed seen in Sri Ramakrishna during his earlier stage of *sadhana*. But no such symptoms are mentioned in the later stages. Possibly acclimatization settles in here also, though it is uncertain whether body-temperature so high as Sri Ramakrishna used to have is possible in sympathetic burst. The physiological subsystems of a saint, passing through many unusual practices so far unexplored by medical science, are indeed little known.

The fact that the subsystems were undergoing changes during sympathetic burst was also established by pole-zero analysis which are shown in Fig-9 and 10. The concept of spectra has already been explained. The frequency where power peaks is called a pole, zero is the frequency where power is zero. A system may have many poles and zeroes. Just like a bridge on a river may be characterized by pillars at specific points and hanging ropes or bars at different angles for support and stability, similarly pole-zero locations specify a system. If the system is steady and stable, poles and zeroes do not change with respect to time. If the system undergoes dynamic change, poles and zeroes move.

Poles and zeroes were obtained in the present case following the

linearization method (Mainardi *et al.* 1995). It may be noted that poles and zeroes are moving considerably in epoch period (Fig. 9), whereas these are almost stationary during non-epoch (Fig.10). The autonomic systems of the subjects were certainly undergoing dynamic changes during epoch periods.

In the late 1920s, Walter Bradford Cannon, the American physiologist, introduced his hypothesis of 'flight and fight' for the cause of the sympathetic burst in the autonomic system. This hypothesis states that when an animal or a man is suddenly confronted with danger, the sense of the need for flight and/or fight triggers the sympathetic system. It causes redistribution of blood (more to heart and skeletal muscles and less to skin and abdominal viscera), increased oxygen intake and higher metabolic rate. These equip the animal/man to fight, or better, run.

In light of this, the possible reason for the triggering of sympathetic activity in Sri Ramakrishna may be as follows: the conscious self is also the protector of the body. It senses, calculates and takes all appropriate measures for the all-round protection and welfare of the organism. During *samadhi*, this body-consciousness merges with the background (ultimate reality). Since the protector dissolves, it is likely that all the physiological subsystems will be alerted. This may trigger the sympathetic activity.

VII. GENESIS OF CONSCIOUSNESS

Ability to know itself, also termed here *chit*, seems to be an inherent property of all elementary particles and it is possibly a quan-tum property. Several arguments are provided below in support of this. The introduction of this quantum state explains the genesis of consciousness avoiding dualism. It also provides a *sense of purpose* to every phenomenon, explains the survival instinct and suggests why the process of evolution is a must in every format of the universe. It also explains why any spatio-temporal structure tries to perpetuate, prefers stability and abhors annihilation. *Chit* seems to be at the core of the anthropic principle.

The following arguments may be cited why *chit* should be a quantum state :

(i) All other properties of elementary particles are quantum in nature, hence *chit* must be so.

(ii) There is an interesting relationship between the depth of

conscious attention and Dt, the estimation of elapsed time. It is a fact that with the increased depth of conscious attention, the accuracy of estimation of elapsed time becomes poorer and poorer. In other words, more focussed conscious attention involves higher energy as is measured in a recent fMRI experiment (Horwitz, Friston and Taylor, 2000). This reminds us of Heisenberg's uncertainty relationship.

(iii) Experience (as the word is used by Chalmers to indicate the 'hard problem' of conscious) in general or thought in particular does not exist independent of particles as photons may. Hence *chit* is not electromagnetic in nature. Thoughts always remain associated with matter as it is in CNS, except its *approximate* symbolic (*e.g.*, linguistic) representation which may be transmitted electro-magnetically.

Quantum properties, almost in every case, were introduced to explain the behaviour of particles. The properties like electron number / muon number / tau number / strangeness were required to take care of the fact that all types of decay are not permissible, but the respective quantum numbers must balance on both sides. The cases of *charge* and *spin* are well known. When *colors* were introduced with the assumption that 'all naturally occurring particles are colorless', many people thought it the last gasp of quantum mechanics (Griffiths, 1987). *Chit* is also introduced here to explain the behaviour of particles and it solves the 'hard problem' of consciousness.

The manifestation of the universe has been possible due to *chit*. Consider an environment where matter is generated continuously and annihilated in the next moment. Within this cosmic play, some *structures* survive and evolve. Which of the two is preferable? It is *chit* which chooses the second. It also explains the absence of large amounts of antimatter.

How is *chit* accumulated? In a charge-like or wave-like (say laser) manner? The second is suggested by some (Jibu *et al.* 1996), but most likely it is executed by a group of cells (Logothetis *et al.* 1989; Crick and Koch 1995) because similar phenomenon are seen in visual and auditory systems. Koch (1995) suggests the location of the thalamus. Anaesthetists suggest that the depth of consciousness is proportional to respiratory sinus arrythmia (Pomfrett 1995) and it is controlled in the medula and brainstem.

In a recent study of brain-computer interface (IEEE Trans, 2000),

m-rhythms, readiness potential, event-related potentials (ERPs) and slow cortical potentials (SCPs) are generated by locked-in patients with volitional movement or just thinking about making a movement. If thought generates SCPs, depolarization is taking place somewhere and movement of ions/neurotransmitters must precede this; in a sense, thought is moving the ions because thought can interact with *chit*.

The best verification of *chit* may come from the study of the particles in specific zones of CNS in the light of Pauli's principle.

REFERENCES

Akselrod, S., D. Gordon, F.A. Ubel, D.C. Shanon, A.C. Barger and R.J. Cohen. 1981. 'Power Spectrum Analysis of Heart Rate Fluctuations: A Quantitative Probe of Beat-to-beat Cardiovascular Control.' *Science*, 213:213.

Bianchi A., B. Bontempi, S. Cerutti, P. Gianoglio, G. Comi and M.G. Natali Sora. 1990. 'Spectral Analysis of Heart Rate Variability Signal and Respiration in Diabetic Subjects.' *Medical and Biological Engineering and Computing*, 29:205.

Billman, G. E., and J.P. Dujardin. 1990. 'Dynamic Changes in Cardiac Vagal Tone as Measured by Times Series Analysis'. *American Journal of Physiology*, 258:H869.

Chalmers, David J. 1996. 'Towards a Science of Consciousness'. *The First Tucson Discussions and Debates*, ed. Stuart R. Hameroff, A.W. Kaszniak and A.C. Scott. Cambridge, Mass.: MIT Press. pp. 5-28.

Crick, F., and C. Koch. 1995. 'Are We Aware of Neural Activity in Primary Visual Cortex?' *Nature*, 375:121-123.

Critchley, M. 1937. 'Musicogenic Epilepsy', *Brain*, 60:13.

Das, S.R., and G.C. Ray. 1998. 'Neural Representation of Music During Musicogenic Epilepsy.' *Proceedings of the Twentieth Annual Int. Conf. of the IEEE Engineering in Medicine and Biology*, Hong Kong, 29th Oct - 1st Nov,20(3/6):1657-60.

Farmer, J.D., E. Ott, and J.A. Yorke. 1983. 'The Dimension of Chaotic Attractors', *Physica*, 7D, 173-180.

Griffiths, David. 1987. *Introduction to Elementary Particles*. London: John Wiley & Sons.

Hortwitz, B., K.J. Friston, and J.G. Taylor. 2000. 'Neural Modelling and Functional Brain Imaging: An Overview.' *Neural Networks*, 13(8-9):829-846.

IEEE Transaction on Rehabilitation Engineering. 2000.8(June) no. 2 (the volume is a special issue on 'brain-computer interface'.

Jibu, M., and K. Yasue. 1993. 'Introduction to Quantum Brain Dynamics' in *Nature, Cognition and System III*, ed. E. Carvallo. London: Kluwer Academic.

Hyndman, B.W., R.I. Kitney and B.M. Sayers. 1971. 'Spontaneous Rhythms in Physiological Control Systems.' *Nature*, 233:339.

Kitney, R.I., and O. Rompelman. 1980. *The Study of Heart Rate Variability*. Oxford: Clarendon Press.

Logothetis, N.K., and J.D. Schall. 1989. 'Neuronal Correlates of Subjective Visual Perception.' *Science*, 245:761-3.

Mainardi, L.T., A.M. Bianchi, G. Baselli, and S. Cerutti. 1995. 'Pole-Zero Tracking Algorithms for the Extraction of time-variant Heart Rate Variability Spectral Parameters.' *IEEE Trans. Biomedical Engineering*, 42(March):250.
Marsden *et al.* 1982. *A Text Book of Epilepsy*, ed. J. Laidlaw and A. Richens. New York: Churchill Livingstone, p.108.
May, R.M. 1976. 'Simple Mathematical Models with Very Complicated Dynamics', *Nature*, 261(June 10):459-67.
Mayer-Kress, G. ed. 1986. 'Dimensions and Entropies in Chaotic Systems.' In *Problems Associated with Dimensional Analysis of Electroencephalogram Data*, by S.P. Layne, G. Mayer-Kress and J. Holzfuss. Berlin: Springer-Verlag.
Merzenich, M.M., P.L. Knight, and G.L. Roth. 1975. 'Representation of Cochlea within the Primary Auditory Cortex in Cat.' *Journal of Neurophysiology*, 28:231-49.
Parker, S.T., and L.O. Chua. 1987. 'Chaos: a Tutorial for Engineers', *Proceedings of IEEE*, 75(8):982-1008.
Poskanzar, D.C., A.E. Brown, and H. Miller. 1962. 'Musicogenic Epilepsy Caused Only By a Discrete Frequency Band of Church Bells'. *Brain*, 85:77.
Prabhavananda, Swami, and Christopher Isherwood. [1944] *Bhagavad Gita*. Chennai: Sri Ramakrishna Math, p.271 (in Appendix).
Ray, G.C. 1995. 'Study of Biological Rhythms During Musicogenic Epilepsy', 17th Annual Int. Conference of the IEEE Engineering in Medicine and Biology, Montreal, Canada, Sept 20 - 23.
———. 1995, 'R-Parameter: a New Measure of Level of Consciousness', *Journal of the Institution of Engineers* (India), 76(September):7-12.
Ray, G.C., and Alexander Y. Kaplan. 1994. 'Trancendent Signal and its Possible Signature on Electroencephalogram.' *Journal of the Institution of Engineers* (India), 74:22-31.
Ray, G.C., Alexander Y. Kaplan, and E. Jovanov. 1997. 'Morphological Variations in ECG During Music-Induced Change in Consciousness', 19th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology, Chicago, Oct 30-Nov 2.
———. 1999. 'Homeostatic Change in the Genesis of ECG During Yogic Breathing', *Journal of the Institution of Engineers* (India), 79(March):28-33.
Rigveda 5.48.5
Rolland, Romain. 1965. *The Life of Ramakrishna*, translated from the French by E.F. Malcolm Smith, Calcutta: Advaita Ashram, p. 23-25.
Sankaracharya. Compiled from *Thus Spake Sri Sankara*, Chennai: Sri Ramakrishna Math, pp. 3, 68, 73, 74, 84 and 86.
Servit, Z., J. Macher, A. Stercova, D. Dudas, M. Kristof and V. Cervenkova. 1962. 'Reflex Influences in the Pathogenesis of Epilepsy in the Light of Clinical Statistics', *Epilepsia*, Amsterdam, 3:315.
Shamma, S.A., J.W. Fleshman, P.W. Wiser, and H. Versuel. 1993. 'Organization of Response Areas in Ferret Primary Cortex', *Journal of Neurophysiology*, 69(2):367-83.
Sri Sri Ramakrishna Kathamrita. 1983. Kolkata: Ananda.
Wang, K., and S.A. Shamma. 1994. 'Self-Normalization and Noise Robustness in Early Auditory Representations'. *IEEE Trans. Speech and Audio Processing*, 2(3):421-35.
———. 1995. 'Spectral Shape Analysis in the Central Auditory System'. *IEEE Trans. Speech and Audio Processing*, 3(5):382-95.

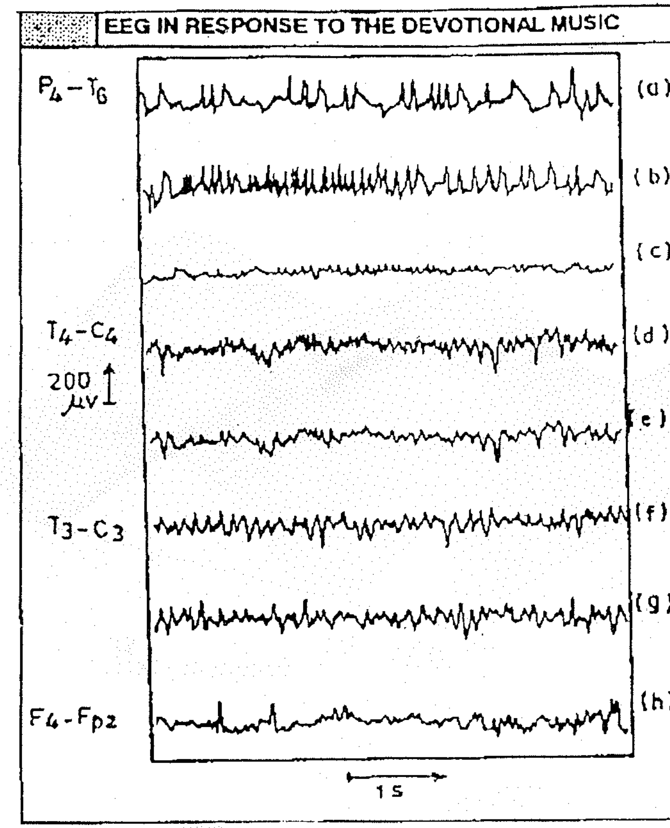


Fig. 1. EEGs showing transcendence

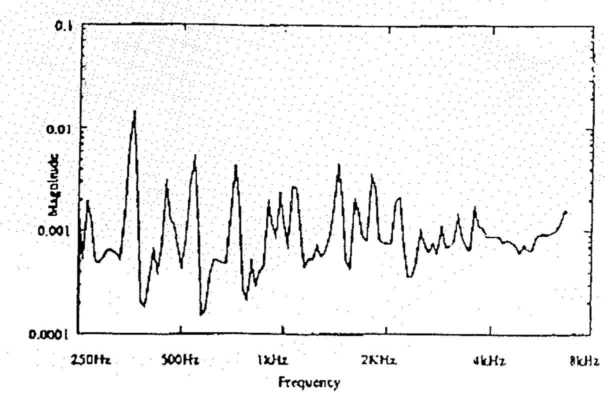


Fig. 2. Auditory spectrum of the sustained rowel/a/

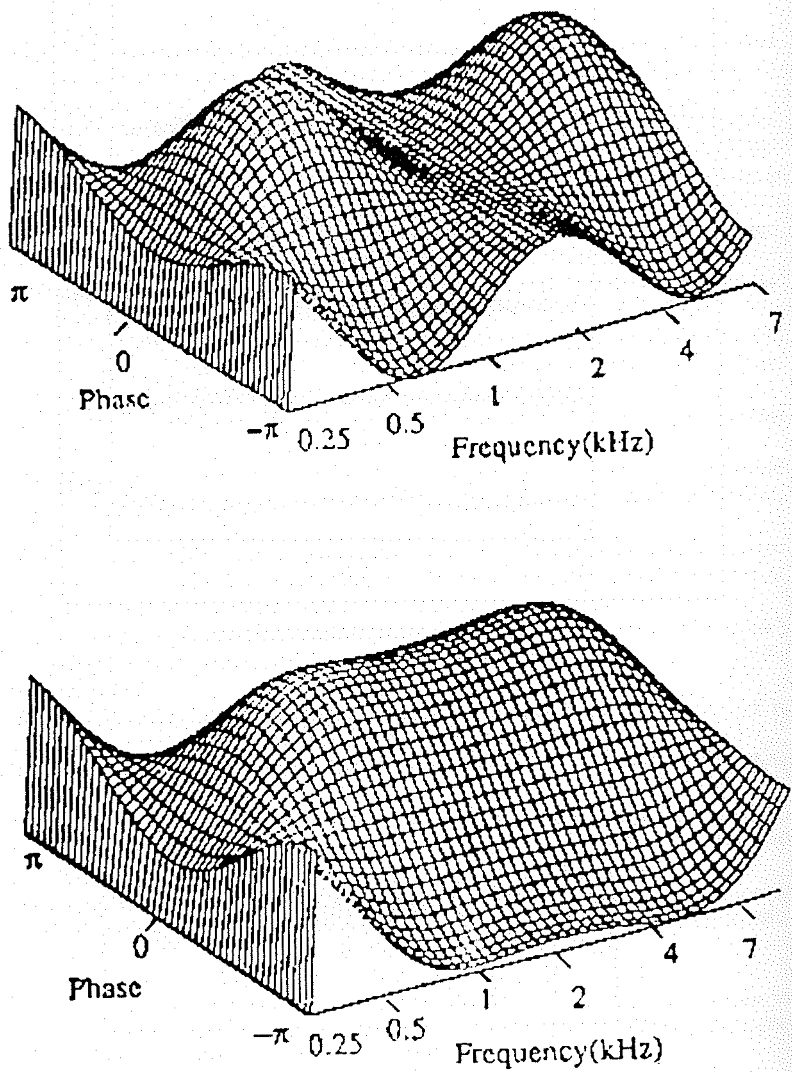


Fig. 3. Cortical representation of music: (a) epoch, (b) non epoch

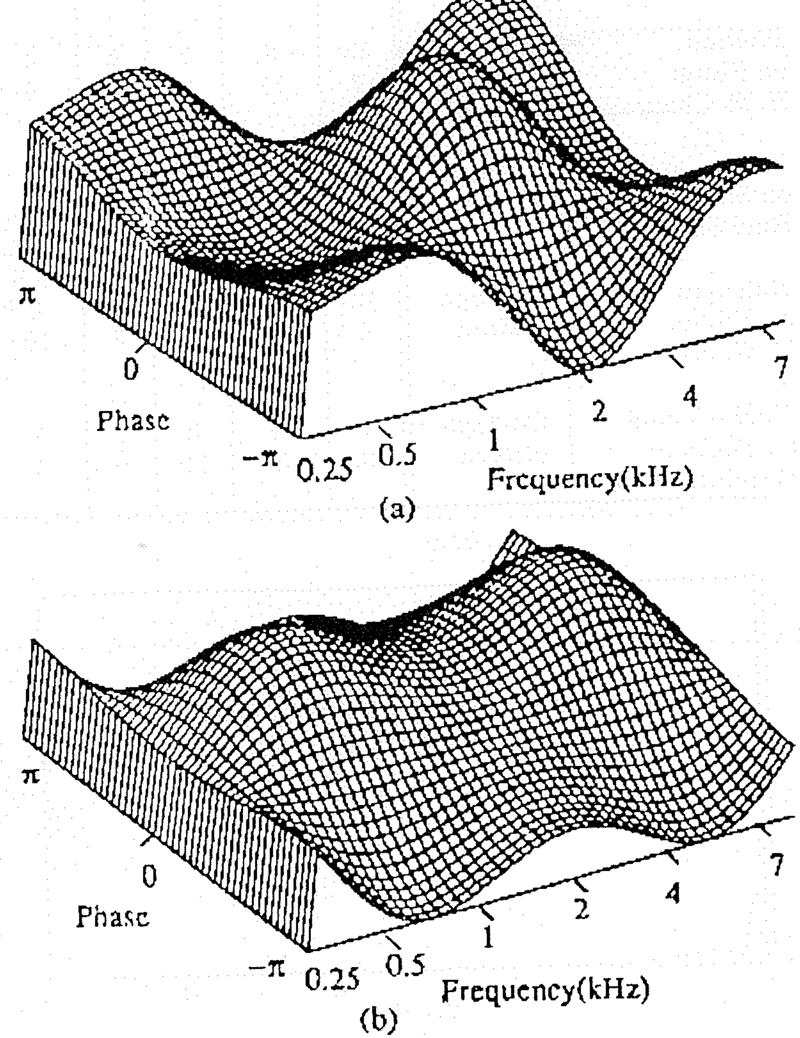


Fig. 4. Cortical representation of (a) monotone, (b) noise

Ragas	Episodes	Binary		Ternary		
		SL	NSL	SL	M	NSL
Jhijhoti on Violin: V. G. Jog	Ch-raga *vistara*	17 6	1 12	14 1	4 7	0 10
Bhatiali on Flute: H. P. Chaurasia	Ch-raga *vistara*	13 7	5 11	7 4	8 10	3 4
Yaman-Manj on Sitar: Ravi Shankar	Ch-raga *vistara*	8 9	10 9	3 7	8 7	7 4
Kalangra on Violin: V. G. Jog	Ch-raga *vistara*	10 6	8 12	6 1	6 7	6 10
Tilak-Kamod on *Shehnai*: Bismillah Khan	Ch-raga *vistara*	7 15	11 3	3 12	7 5	8 1

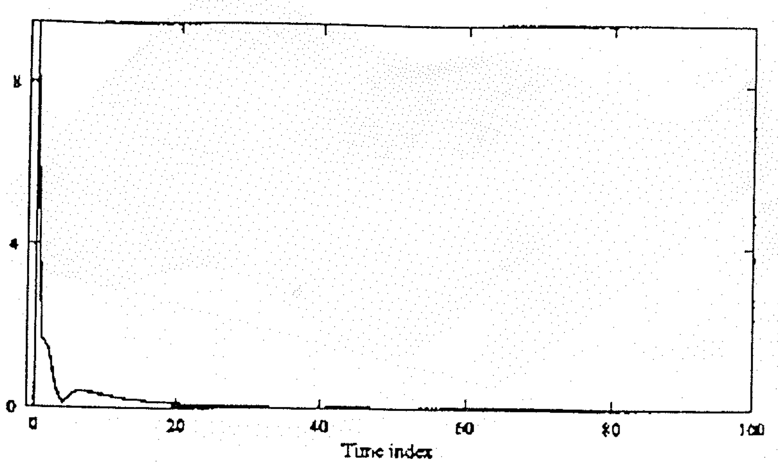


Fig. 5. Inverse transformer of ripple in time domain

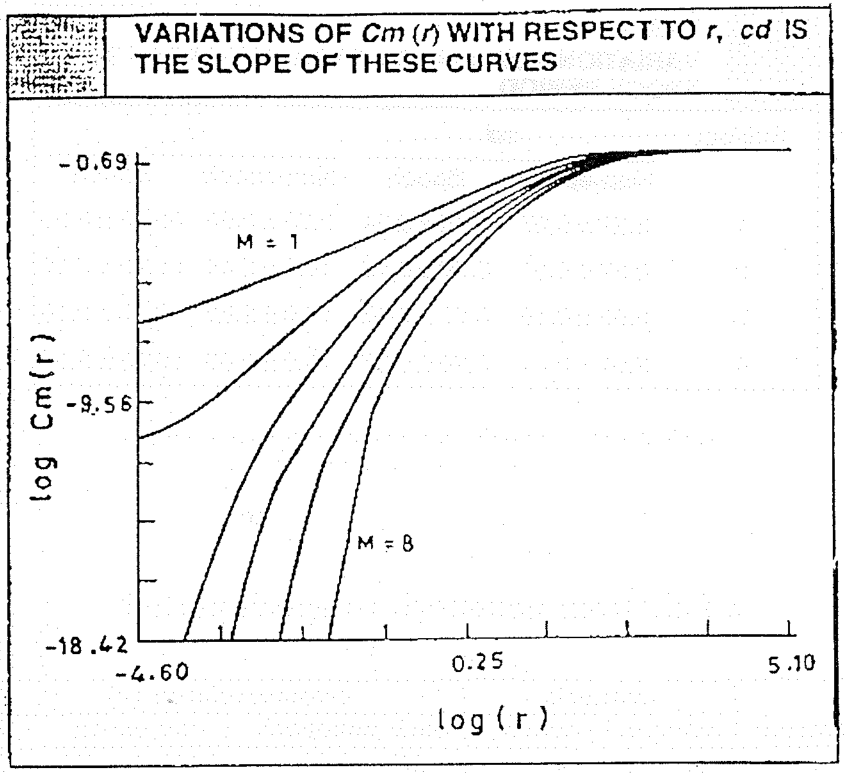


Fig. 6. Correlation dimension in the slope of the curves

TABLE
VARIATIONS OF *cd* AND *r* FROM NON-EPOCH TO EPOCH PERIOD

Subject	*cd*		*r*	
	Non-epoch	Epoch	Non-epoch	Epoch
1	3.27 ± 0.27	2.48 ± 0.24	9.33 ± 3.68	05.83 ± 3.29
2	3.57 ± 0.37	3.00 ± 0.13	10.0 ± 2.44	08.25 ± 2.16
3	3.58 ± 0.18	3.07 ± 0.17	14.0 ± 2.83	10.75 ± 2.16
4	3.64 ± 0.61	2.90 ± 0.37	12.0 ± 0.00	10.33 ± 2.36

Table 2. Correlation dimension and r-parameter reduces in the epoch period

ρ falls from nonepoch to epoch period

session	channels		crosscorrelaton (ρ)	
	nonepoch	epoch	both nonepoch	one epoch
session-1	C4 - T4	P4 - T6	0.277 ± 0.090	0.007 ± 0.050
session-2	F4 - C4	C4 - P4	0.503 ± 0.045	0.460 ± 0.054
session-3	P3 - O1	C3 - P3	0.447 ± 0.058	0.420 ± 0.065
session-4	T6 - O2	T4 - T6	-0.021 ± 0.116	-0.195 ± 0.091

Table 3. Cross-relation between different parts of the train during epoch and non-epoch

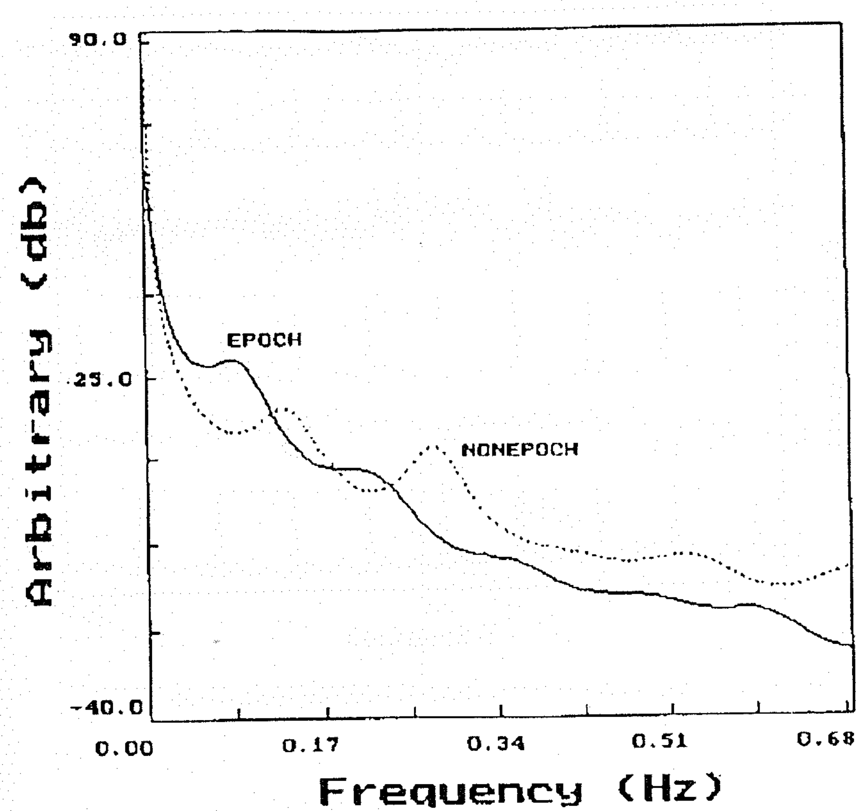


Fig. 7. The HRV Spectrum

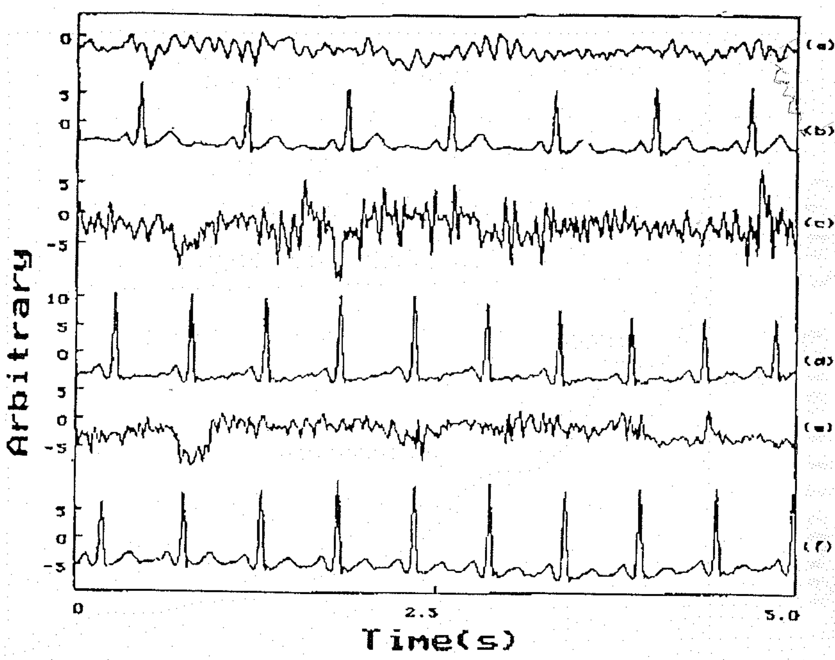


Fig. 8. EEGs and ECGs during non-epoch and epoch periods

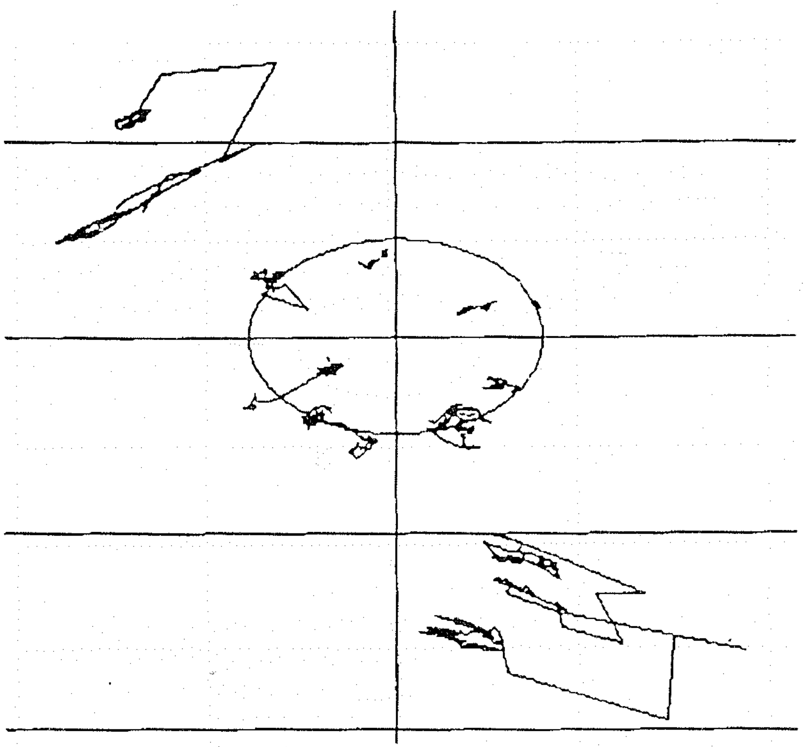


Fig. 9. Pole-zero locations and their movement during the epoch period

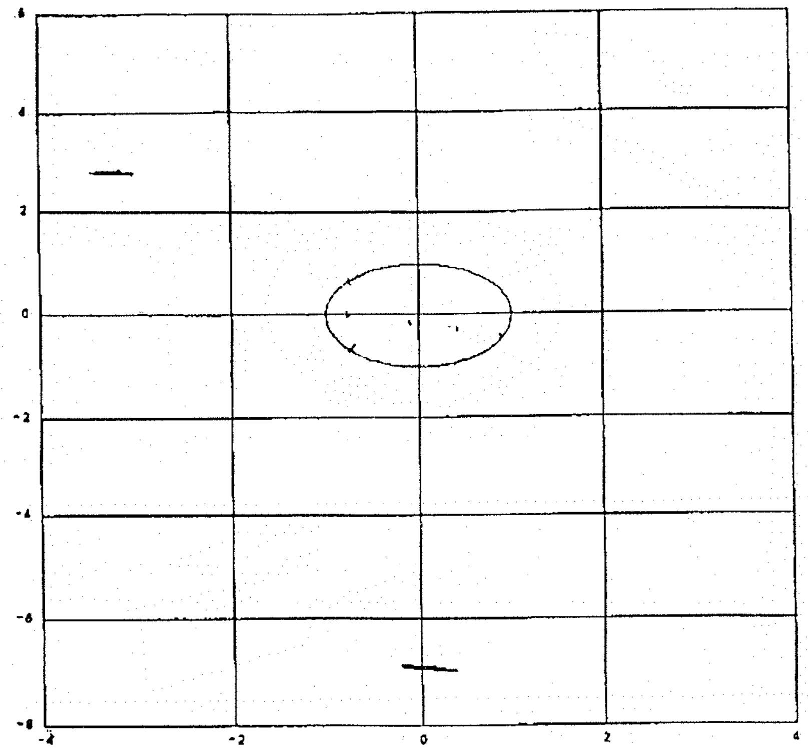


Fig. 10. Pole-zero locations are stationary during the non-epoch periods

Neuropsychology of Consciousness

SHOBINI L. RAO

Introduction

The ubiquitous nature of consciousness begs its serious study. The questions that arise in the mind are regarding its nature i.e. what is consciousness? its location i.e. where is consciousness? Recent scientific studies have converged on the view that consciousness is a process and not a substance. The philosophical debate between monistic reductionism and dualism have been faced with evidence from cognitive neuroscience which has amply shown that mental processes which are conscious require both the brain and the environment. The interaction between brain and the environment give rise to the mental processes including consciousness.

Nature of Mental Processes from the Cognitive Neuroscience Perspective

Cognitive Neuroscience has utilised the Lesion technique and the Imaging techniques to understand the Brain Behaviour relationships. These studies have made it abundantly clear that mental processes arise from an interaction of the brain with the environment. Alterations of the environment changes behaviour. The environment may be nutritional, physical, or psychosocial. Malnourishment is associated with poor intellectual functioning. Imbibing intoxicating substances such as alcohol alters a wide range of mental processes such as attention, memory and emotional functioning (Schuckit, 1995). The role of psychosocial environment in shaping behaviour is very well established in the fields of Clinical Psychology and Psychiatry. Opportunities presented by the environment for intellectual stimulation, for learning different skills such as reading or writing, for emotional solace and comfort shape the child's growth and development in multiple fields. The growth of attention, intelligence, language are dependent on the environment in crucial ways. The development of joint attention which is the precursor for social interaction is dependent on the availability of the mother /

care giver to the child. Language requires the hearing of natural languages. Shaping of utterances of the child by the repeated feedback from the adults in the environment teaches the child to speak in the mother tongue or another language. The role of the environment is powerful and yet subtle in the realm of emotions. Expression and comprehension of emotions are influenced by the environment. Excessive criticisms or a high degree of negative emotions are one of the etiological reasons for the serious mental disorder called schizophrenia. The lack of family support as well as social support from friends maintain depression and retard a cure from it. In fact the role of the environment in shaping mental processes such as our thoughts and feelings is clear and evident. However it is difficult to conceptualise the role of the brain.

The brain may be thought of as the medium in which the influences of the environment act to give rise to mental processes. The nature of this action of the environment on the brain is in the form of an interaction rather than in the form of an agent and an effect. This interaction can be understood if the unique nature of the brain is understood. The brain is an organ of the body which is influenced by the environment and which is also able to influence the environment. The other organs of the body are influenced by the environment, but are unable to independently change the environment. The brain has the capacity to alter the environment. This two pronged interaction with the environment is possible because of the dynamic and open nature of brain functioning. Approximately the brain has a billion neurons with a trillion connections. Each neuron is capable of processing information. The high density of neurons and the high degree of connectivity make it possible for the brain to process the enormous amount of information that we face all the time. The brain is not static. It undergoes changes in its chemical milieu, metabolic milieu, and in the electrical activity of its different regions every moment. This is in response to the environmental demands as well as part of the housekeeping functions. (Kandel, Schwartz & Jessell 2000). What is more interesting to the subject of consciousness is its capacity to undergo changes in connectivity in response to the environmental demands.

The brain has several important capacities from the perspective of understanding mental processes. The first is its capacity to process information. It is this capacity which enables us to sense and perceive the environment. The second capacity is to store the information that it has processed. The storing of information is the basis of memory. The third capacity is the brain's capacity to learn. Learning is essential to link isolated bits of information to make a bigger picture. Learning and memory are closely interdependent. The buzz of external world makes sense and mental processes arise because of the these primary processes. Thus the bedrock of our mental life are information processing, learning and memory. Learning and memory are dependent on the brain's plasticity. Plasticity may be defined as the brain's capacity to change in response to environmental demands. This capacity is the basis of learning, maturation, growth and recovery of lost functions. Growth and development of the child's mental capacities occurs because of the brain's plasticity. The child and the adult are able to learn because of the brain's plasticity. The injured brain is able to heal itself again because of the plasticity.

Brain plasticity is mediated by several neural events. Structural changes as in the formation of new connections or pruning of old connections is an important means of enabling plasticity. Functional modifications arising from denervation supersensitivity in the dendrites of the damaged neurones; disinhibition leading to responsiveness in previously silent synapses; are some of the other means of plasticity. The intact adult brain shows plasticity in the form of learning. In fact learning is the most elegant, simple and powerful demonstration of brain plasticity. It is in the process of learning that the brain's responsiveness to the environment is crystal clear.

The processes of learning and memory are the basis for how changes take place in the brain's capacity to make sense of the environment. At any moment of time current sensations are understood in the background of previous experience stored as memories to form perceptions. Sensations make meaningful perceptions because of our previous memories. Thus learning which makes material for memory and memory by itself are the means by which the brain is able to transform itself.

Transformation of the brain through experience

Experience arises out of the process of learning and memory. Memory transforms the manner in which the brain perceives the environment. Learning a new language offers a good example of this capacity of the brain. The new language is a buzz of sounds

when it is first heard. Learning to break the sounds into meaningful bits, and learning to link the sound to meaning makes the language meaningful. The initial stand alone sounds become blended sounds which become words. The experienced speaker identifies words rather than individual sounds which are later constructed to blended words or individual words. The brain is transformed to process words rather than construct words from individual sounds. How is this transformation achieved? The answer lies in establishing functional connectivity or connecting discrete brain areas which process different aspects of the information needed to recognise words when they are heard. A simple account of this is that the auditory cortices which distinguish sounds are connected to the phonemic processing areas in the posterior superior temporal lobes. These are in turn connected to the angular gyri in the left parietal cortex to access meaning. Perceptions of words are also influenced by inferior prefrontal areas in the left prefrontal lobe which mediates semantic memory. Establishing this connectivity is at the core of learning a language which converts the initial buzz of sounds to a meaningful sound pattern. The connectivity and the consequent learning changes or transforms the way we perceive the environment

Another agent of transforming the way in which the brain interacts with the environment is attention. Attention shapes our perception, influences our learning and determines the recall from memory. Alertness is the basis of attention. A sleeping person cannot be attentive. Attention may be defined as channelled alertness. Attention chooses the aspects of the environment which are perceived and acted upon. It also sustains the alertness for continued processing of information. This focusing and sustenance of attention determines to a large extent what the brain is processing at any given moment of time. Attention selects the information which is later bound together by learning and stored by memory. Thus attention, learning and memory can be termed the three skills by which the brain is able to fill the experiential world with the rich content which form our thoughts and feelings. Attention transforms the experience in that what is attended is experienced and what is ignored is not.

Characteristics of attention

1. Attention is a quick process compared with learning. Learning requires repeated exposures to the same context of stimuli, while attention is mercurial in what it chooses to be under its searchlight.

2. Attention is not fixed in the space time continuum. It is the ability of attention to traverse the space time continuum which enables seamless blending of the past with the present. Attention may be directed at the sensations of the present moment or the memories of the past moments.

3. Attention does not blur the space time divisions. Attention is shifted to discrete points in space. Attention can be directed to the present moment. The present moment can be ignored and attention can be directed to a past memory.

4. Attention is multidimensional. Attention is modality specific and supra modal. The posterior regions of the brain mediate modality specific attention. This is termed as the posterior attention system. These attention systems which have been studied in detail for the visual and auditory modalities. The primary role of the posterior attention system is to amplify sensory processing through attention amplification of sensory processing. The role of the anterior attention system is to decide between alternatives, to suppress unwanted or irrelevant responses; to channel volitional processes; and to maintain the alertness levels. The anterior cingulate and the orbito frontal cortices are crucial for the anterior attention system (Posner & Dehaene 2000).

5. Attention can also be fixed to a context of stimuli or to a certain stimulus. Motivational influences, emotional evaluations, stimulus characteristics such as size and colour can fixate attention to a certain stimulus.

6. The rapidity with which attention can shift in time and space gives it a quality of continuity.

7. Attention has an object to attend. Attention when it exists without an object is alertness.

Attention and consciousness

Is attention identical with consciousness? Consciousness also requires alertness, is able to shift rapidly across space and time, has a sense of continuity, and has space and time divisions. However Consciousness has a sense of agency which attention does not. The "I" of psychology or "ego" "*ahamkara*" of the Indian philosophy is characteristic of consciousness and not that of attention.

The agency or *ahamkara* differentiates consciousness from

attention. If it is this one difference alone why is it so easy to study attention scientifically. The psychological characteristics of attention and the brain mechanisms of attention are fairly well understood. Why is it so difficult to find out the where and how of consciousness. The answer appears to lie in our mistake of mystifying consciousness. The rapid strides of understanding mediation of mental processes by the brain has arisen from a componential analysis of mental processes. This is followed by attempts to identify the brain bases of each of the components. Applying a similar analysis yields discrete processes of consciousness which can then be localised to either discrete brain regions or to discrete brain processes. Conceptually it is a doable exercise (Rao 1999). Direct experimental evidence is available for some of the concepts. The neural basis for awareness has been experimentally demonstrated in studies on sometosensory perception and vision. Direct stimulation of the somatosensory cortex, ventrobasal thalamus with pulse trains at liminal intensity found that the stimulation had to be present for at least 500 milliseconds for the sensation to become conscious. Stimuli below that intensity or duration did not become conscious (Libet 1979). Horizontal or vertical gratings were seperately presented to the two eyes of alert monkeys. The animals had been trained to report which direction of the grating it saw. A distinct group of neurons in the area MT fired whenever there was an upward direction of movement indicating an awareness of the direction of movement. Other neurons in MT fired in response to motion in either direction (Logothetis & Schall 1989). Binding of different neuronal activations seems to be necessary for building a dominant focus of awareness. Anosognosia is a clinical condition wherein the patient does not recognise the diseased body part. A dissociation between meaning and awareness was seen in these patients. The patient would deny the impairment verbally but would not attempt to use the limb. Again the patient may verbally accept the impairment but would attempt to use it (Bisiach & Gemianini 1991). Yet other components of consciousness remain at the conceptual stage itself. The continuity in awareness is explained by the theory of dominant focus. Multiple drafts of the experience and the environment are being made in succession. Each draft becomes the content of a dominant focus. Consciousness would be experienced as a stream (Dennett & Kinsbourne 1992)

Componential analysis of agency

The sense of agency is an important ingredient of consciousness. It is often thought to be synonymous with consciousness. Agency is the sense of "I" or "*ahamkara*". The author had previously proposed that interaction with a physical environment which has a space-time Continuum requires a sense of agency (Rao, 2001). The *ahamkara* or "I" is necessary to divide experience in terms of this space and time as well as to integrate experience in terms of space and time. Space and time are processed in the brain by discrete areas. The important brain areas which process space are the parietal lobes and the hippocampus. Lesions in parietal areas in man are associated with difficulties in processing visuo-spatial information as well as in the processing of spatial information. Form perception, localisation of objects in space, size and shape constancy are affected. Route finding difficulties are also evident in patients with parietal lesions (Joseph 1996) Hippocampal lesions particularly of the right side is associated with difficulties in the processing of allocentric and ego-centric space. Patients have difficulty in spatial localisation of objects if the orientation of their body is changed (Abrahams et. al. 1997). Hippocampus is the centre for spatial cognition in animals.

The processing of time is mediated by different brain areas depending on the time scale. The prefrontal areas mediate the computation of time in the milliseconds range. The cerebellum computes time in the microsecond range (Mangels et. al. 1998). It is herewith hypothesised that the computation of time and space are not isolated from each other. There are connections between the parietal lobe and the hippocampus responsible for the computation of space with the prefrontal areas responsible for the computation of time. There are connections between the cerebellum responsible for the computation of time with the parietal lobes responsible for the computation of space. The further postulate is that divisions of space and time require an anchor. The present is with reference to the past and the future. Anchoring of time occurs with reference to the present moment. Anchoring of space occurs with reference to the positioning of the body in that particular moment. Change in the body position has to be computed with reference to that in the previous moment. Therefore computation of space and time which is essential for our survival in the physical environment, requires an anchor for the present moment and the

current body position. The absence of this agency would result in a sense of discontinuity of experience. Memory together with agency give the continuity of experience in this space-time continuum.

The role of attention in consciousness

The components of consciousness are:

1. Alertness
2. Division of experience into the space - time continuum
3. Integration of experience in terms of the space-time continuum
4. The sense of agency to enable the processes in 2 and 3 above.
5. Attention to choose the contents for perception and the contents from memory.
6. The process of perception and recall from memory to supply the contents.

It is seen from the above rather rudimentary list of the components of consciousness that attention is an important component of consciousness, but it is not synonymous with it. The important elements of consciousness are:

1. Alertness
2. Attention
3. The space-time division
4. The space-time integration
5. The anchor of the self or "I".

The specific contribution of attention to consciousness seems to be the choice of contents for consciousness. Attention focuses on specific stimuli or specific memories. Through this selective focusing attention ensures that the contents of consciousness are according to the motivational goals of the organism. The analogy that comes to the mind is that of a sieve which filters the irrelevant stimuli or memories out to supply only the relevant contents to consciousness. Attention has been described as a limited capacity mechanism which acts a filter to select stimuli for further processing (Broadbent, 1971). The question arises: does attention only choose contents for consciousness or does it change consciousness in any other manner? The answer to this question comes from observations on patients with psychiatric illness, patients with brain damage, and persons who practice meditation.

Mental illness, attention and consciousness

Schizophrenia is a mental illness which is associated with disturbances of attention. An inefficient attention filter which is not able to filter out irrelevant information is one of the core psychological deficits in schizophrenia (Siegel, 1995). The illness is characterised by gross disturbances in consciousness. The patient has bizzare experiences such as auditory hallucinations, poor reality testing, emotional blunting, strange bodily sensations. There is social isolation, alienation and poor social as well occupational functioning. Neuropsychological assessment has shown a spectrum of deficits in the realms of attention, executive functions and memory. The important point to note is the deficit in executive functions. Executive functions are the functions which enable goal directed behaviour. Examples of executive functions are allocation of attention resource or mental effort, working memory, planning, set shifting, abstraction, concept formation. The executive functions manipulate and control the nature of processing of the perceptual and mnemonic information in order to achieve the goals of the organism. In schizophrenia disturbances of executive functions are closely associated with the disturbances of consciousness which characterise this illness. The patient has a schism in the psyche, with a fragmentation of consciousness. The chief deficit is in attention, be it in the focusing of attention or in sustaining of attention.

Another example of disturbances in consciousness associated with disturbances in attention is from attention deficit hyperactive disorder. Children with ADHD are impulsive, easily distracted and unable to sustain attention even for brief periods of time. The child is unable to learn normally because of the distractibility. Orbito frontal damage has been implicated by the pattern of neuropsychological deficits (Taylor, 1994). Phenomenologically the consciousness of the child appears to have lost the continuity and stream like quality. The child is jumping from one task to another without any apparent purpose. Attention deficit is closely associated with this segmentation of consciousness.

The third example of disturbances in consciousness associated with deficits of attention is from autism. Autism is a pervasive developmental disorder. The child has stereotyped movements, poor language skills and is unable to form emotional and social relationships. Joint attention which develops from about 8 months

of age is poorly developed in these children. Joint attention is characterised by pointing to objects; looking in the direction of an object pointed to by an adult; being able to direct attention towards different aspects of the environment depending on the behaviour of others. The autistic child has deficits in this joint attention (Lord & Rutter, 1994). The consciousness of the child is characterised by a rigidity in behaviour. The child is unable to shift attention from the object or context which holds the attention. There is a persistence in the attention and a rigidity in the consciousness.

The lack of selectivity of attention in schizophrenia fragments consciousness. The distractibility of attention in ADHD segments consciousness. The inability to orient attention to changing events such as social cues restricts consciousness. Thus attention conditions consciousness in terms of its range, nature and responsiveness. Attention does not merely supply the contents of consciousness. Further proof of the influence of attention on consciousness comes from the field of studies on meditation.

Meditation, attention and consciousness

Research has tried to unravel the effects of mediation on brain functioning. Meditation is a technique practised for mind control world wide. The two types of meditation are focused concentration or attention on an object and relaxed mindfulness. The yogic path of meditation with the three staged approach of *dharana* or drawing of the attention towards an object of contemplation, *dhyana* or sustained attention to the object of concentration and *samadhi* or merging of attention with the object of contemplation, belongs to the path of focused attention. The Buddhist traditions of Zen and Vipassana schools practice mindful attentiveness. Here there is no one object of concentration. Attention is channelled in both types of meditation. Meditation is accompanied by subjective reports of alteration of consciousness. Calmness, serenity, peace and bliss are the subjective experience of meditators. The commonality of these experiences across religions and cultures denote a physiological basis for the same. The quest to understand the brain mechanisms mediating these experiences got a thrust with the advent of functional imaging technologies.

Meditation was associated with a significant increase of rCBF in the cingulate gyrus, inferior and orbital frontal cortex, dorsolateral prefrontal cortex and thalamus. There was a significant negative correlation between the activation in left dorsolateral prefrontal cortex with that in the left superior parietal lobe. This correlation has been interpreted as reflecting an altered sense of space experienced during meditation. The increased rCBF in the frontal areas has been interpreted as reflecting focused concentration. Thalamic increases are hypothesised to indicate overall increased activity during meditation. The rCBF pattern in meditation is distinct from rest and is complex in nature (Newberg et. al. 2001). Brain mapping of event-related electrical activity has shown that the frontal theta rhythm is active during meditation (Kubota et. al. 2001). The areas which are active during meditation are different from those that are active during hypnosis, sleep and even simple rest (Rao' 2001a). Meditation is thereby an active process and not a passive process. Further, experienced meditators reported periods of blissful experiences while they were meditating. The number of episodes of blissful experiences was correlated with the different parameters of the EEG (Caftans & Golocheikine 2001).

Meditation is essentially tuning in of attention to one object (focused attention) or the tuning out of attention from being fixed on any one object (mindfulness meditation). Attention is trained in either direction. The consequence is not only changes in the contents of consciousness but also changes in the quality of consciousness. Subjective reports of Indian seers who have practised meditation for long periods of time report a loss of the "*ahamkara*" or "I" sense; a feeling of expansiveness; blissful experiences as the outcomes (Svami Turiyananda 1991). These texts speak of a distinct alteration of consciousness due to changing attention. It is not merely the contents of consciousness which is changed but it is the quality and nature of consciousness itself.

Attention and consciousness

Mental illness is associated with constrains on attention. The constraints are in terms of deficits in specific aspects of attention such as the focusing of attention as in Schizophrenia; sustenance of attention as in ADHD; shifting of attention as in autism. Consciousness is deranged in specifc ways in association with the deficits of attention. The deficits of consciousness are in terms of fragmentation, segmentation, rigidity respectively. On the other hand the

tuning in of attention and training of attention as in meditation is associated with alterations of consciousness in the positive direction. Quietude, bliss, expansiveness are the qualitative changes in consciousness. Thus it is seen that whether the alteration of attention is in the positive direction or in the negative direction there is a change in the quality of consciousness itself. The change is beyond the mere change of the contents of consciousness. What could be the mechanism through which this is achieved?

Focusing of attention in a changing environment helps to strengthen the sense of agency. Only when the attention is focused on changing events or stimuli is a sense of space and time created. This is because the shifting of attention makes it possible to process different stimuli which are presented in a changing environment. The presence of stimuli in different places on the physical environment or at different points of time create the mental representation of time and space. The anchor for this division is the sense of self or "I" or "*ahamkara*". Thus attention strengthens the anchor or "I" consciousness or agency. This interaction between the sense of self and attention shapes consciousness. Mental illness is associated with alterations in consciousness, alterations in the sense of agency and disturbances in attention. Meditation is also associated with alterations in consciousness, alterations in the sense of agency and disturbances of attention. The alterations of consciousness in mental illness are in the nature of fragmentation, segmentation and rigidity. The alterations of consciousness in meditation are in the nature of quietness, bliss, fusion and elimination of the sense of agency. The jump from *savikalpa samadhi* to *nirvikalpa samadhi* is described in detail in the vedantic and yogic tradition. *Savikalpa samadhi* refers to one pointed consciousness which is differentiated. The differentiation is in terms of the knower, the knowing and the known. The knower is the "I" or agency, knowing refers to the process of knowing, known refers to the object of the knowledge. After attaining *savikalpa samadhi,* further practice of meditation or tuning of attention leads to *nirvikalpa samadhi.* Here consciousness is not differentiated. There is a merger of the knower, known and knowing. The point to note is that practice of attention tuning through meditation is the technique for achieving this nature of consciousness. The undifferentiated consciousness is also called as the *turiya* state (Svami Turiyananda 1991). Attention which differentiates consciousness by making the changing world perceptible is also the tool by which change is no longer perceived. Consciousness becomes undifferentiated because the space time dimension collapses. Consequently there are no objects to be known or the knower and the process of knowing. Thus attention and the space time differentiation which it enables creates a phenomenal world of space and time. The same attention when it becomes completely focused ceases to differentiate the space and time and consciousness becomes alert but undifferentiated. The phenomenal world that we perceive ceases to exist in this state. Thus attention becomes the creator of the phenomenal world and the sense of agency as well as the destroyer of the phenomenal world and the sense of agency. Consciousness is conditioned by attention. Attention gives both structure and contents to consciousness.

REFERENCES

Abrahams, S., A. Pickering, C. E. Polkey & R. G. Morris (1997) Spatial memory deficits in patients with unilateral damage to the right hipocampal formation. *Neuropsychologia,* 35, 11-24.

Bisiach, E. & G. Gemianini (1991) Anosognsia related to hemiplegia and hemianopia. In G.P. Prigatano & D. L Schachter,New York: Oxford University Press.

Broadbent, D. E. (1971) *Decision and Stress.* New York: Academic Press.

Caftans, L. I. & S. A. Golocheikine (2001) Human anterior and frontal midline theta and lower alpha reflect emotionally positive state and internalised attention: high resolution EEG investigation of meditation. *Neuroscience Letters,* 310, 57-60.

Dennet, D. C. & M. Kinsbourne (1992) Time and the Observer. The where and when on consciousness in the brain. *Behavioural and Brain Sciences,* 15, 183-247.

Joseph, R. (1996) *Neuropsychiatry, Neuropsychology and Clinical Neuroscience,* Baltimore: Williams & Wilkins.

Kandel, E.R., J. H. Schwartz & T. M. Jessell (2000) Principles of Neural Science, New York: McGraw Hill.

Kubota, Y., W. Sato, M. Toichi, T. Murai, T. Okada, A. Hayashi, A. Senguko (2001) Frontal midline theta rhythm is correlated with cardiac autonomic activities during the performance of an attention demanding meditation procedure. *Cognitive Brain Research,* 11, 281-287.

Libet, B. (1979) Subjective referral of the timing for a conscious sencory experience: A functional role for the somatosensory specific projection system in man. *Brain,* 102, 191-222.

Logothetis, N. & J. Schall (1989) Neuronal correlates of subjective visual peception. *Science,* 245, 761-763.

Lord, C. & M. Rutter (1994) Autism and Pervasive Developmental Disorders. In M. Rutter, E. Taylor & L. Hersov (eds.) *Child and Adolescent Psychiatry,* London: Blackwell Science, 569-593.

Mangels, J. A., R. B. Ivry, N. Shimizu (1998) Dissociable contributions of the prefrontal and neocerebellar cortex to time perception. *Cognitive Brain Research*, 7, 15-39.

Newberg, A., A. Alive, M. Baime, M. Pourdehand, J. Santana, E. d'Aquili (2001) The measurement of regional cerebral blood flow during the complex cognitive task of meditation: a preliminary SPECT study. *Psychiatry Research*, 106, 113-122.

Posner, M. I. & S. Dehaene (2000) attentional Networks. In M. S. Gazzaniga (ed.) *Cognitive Neuroscience*, Oxford: Blackwell, 156-164.

Rao, S. L. (1999) The Neural basis of the Attributes of Consciousness. In S. Menon, M. G. Narasimhan, A. Sinha, B. V. Sreekantan *Scientific and Philosophical Studies on Consciousness*, Bangalore: National Institute of Advanced Studies, 114-126.

——. (2001) Meditation and the Brain, Paper presented at the International Conference on Yoga, Swami vivekananda Center for .... Bangalore,

——. (2001) The Elusive Self in the Spatio Temporal Net Paper presented at the International Conference on Science and Metaphysics, National Institute of Advanced Study, Bangalore, June 2001

——. (2002) The elusive self caught in the spatio temporal net. National Institute of Advanced Studies (In Press).

——. (2001) Meditation and the Brain, Paper Presented at the Seminar on Neural Plasticity and Yoga, sVYASA Prashanti Kuteeram, Jigani, Bangalore, November 2001,

Schuckit, M. A. (1995) Alcohol Related Disorders. In H. I. Kaplan & B. J. Sadock (eds.) *Comprehensive Textbook of Psychiatry / VI Vol 1*, Baltimore: Williams & Wilkins, 775-791.

Siegel, D. J. (1995) Perception and Cognition. In H. I. Kaplan & B. J. Sadock (ed.) *Comprehensive Textbook of Psychiatry / VI Vol. 1*, Baltimore: Williams & Wilkins, 277-291.

Svami Turiyananda, (1991) *VivekaCudamani of Sri Sankaracharya*, Madras: Sri Ramakrishna Math.

Taylor, E. (1994) Syndromes of attention Deficit and Overactivity. In M. Rutter, E. Taylor, & L. Hersov (ed.) *Child and Adolescent Psychiatry*, London: Blackwell Science, 285-307.

Co-operation in Apes and Humans

INGAR BRINCK and PETER GÄRDENFORS

1. Co-operation

Human consciousness serves many roles. It helps us imagine, dream, and think rationally. During most conscious activities, the person involved is self-aware. But consciousness is not only of oneself and the external world, but also of other agents and their minds. In this article, we will focus on the interplay between the capacities to co-operate and represent the minds of others. Full-blown co-operation requires not only being conscious of oneself and reflecting on one's own thoughts, but also being conscious of the minds of other agents.

Both human beings and animals co-operate in order to reach common goals. There are many ways of co-operating. Some of these ways may not merit being called co-operation in the literal sense of the word, as it is used about humans. Among these one may count instinctive co-ordination of behaviour, such as it emerges among ants building heaps or honeybees gathering food. At the opposite side of the scale, we find co-operation that builds on elaborate, long-term planning and discussion among the subjects who co-operate. Our aim is to compare and elucidate the similarities and differences between humans and apes as concerns co-operation. There is no doubt that apes co-operate, but, as we shall argue, humans are able to do so in more flexible ways.

Co-operating means working together for a joint benefit: agents together employ a certain means, or series of actions, to achieve a common end, or goal. Co-operation can be achieved directly by co-ordination of behaviour. It can also arise indirectly through a mutual sharing of representations of means and end.

In many cases, it seems that representations are not needed to stand in for the real thing, that is, for the means and the goal. If the co-operative goal is present in the actual context, for instance, water to drink, food to be had, or an antagonist to fight, it is not necessary to focus on a joint representation of it before taking action.

But if, on the other hand, the goal is distant in time or space,

then a common representation of it must be produced before action can be taken. Humans have a powerful tool in language when it comes to creating common goals. Language, as opposed to animal signalling, is not bound by the actual. It is based on the use of representations as stand-ins for actual entities. Use of representations replaces the use of environmental features in communication.

2. *Competitive versus collaborative co-operation*

Co-operation is often competitive. This means that it is undertaken with the view to out-compete, or win over, others. Co-operation may arise in direct competition about resources or as a response to threat. A group of apes may co-operate in order to defend themselves against a predator, or in order to get hold of some food from which the whole group will benefit. In both cases, the group will normally be better off working together than if a solitary ape would have taken action. A group may also compete directly with another group for resources.

But it may also be the case that the individuals that the competitive co-operation is directed at are not involved in the actual co-operation. For instance, one way to take the advantage over a particular rival is by forming an alliance or coalition with other individuals, as an intermediary to the ultimate goal of winning over the rival. In that case, an agent may form an alliance with another agent in relation to an immediate goal. This is a case of short-term planning. However, the ultimate goal of the first agent is to achieve an advantage over a third party, a case of long-term planning. The following is an example of how apes may engage in alliances.

The zoologist Frans de Waal, in his classic book *Chimpanzee Politics*, describes how clever high-ranking chimpanzees are at manipulating others. The story of Yeroen is typical. He had for a long time been the dominant male in the flock, but was dislodged by Luik and later defeated by the young Nikkie. Instead of retiring in bitterness, Yeroen now formed an alliance with Nikkie. Together they defeated Luik, who had sole right to the females. With the new ranking order – Nikkie at the top and Yeroen in second place – Yeroen could use his position to mate with some of the females. Nikkie could not protest with any vigour against this, since he was dependent on Yeroen's support in the struggle against Luik.

Co-operation presupposes that each participant understands what the other participants' actions aim at. It seems that actions become understandable among apes when the actions appear in a competitive context, whether the competition is within or between groups. One ape understands why another ape performs a certain action, because the action is performed against a background of given and limited resources – food, partners, et cetera. The variables that are relevant in a certain situation are given by the context, and by having access to these, the ape can interpret the action accordingly. This line of thought is useful to explain another oftmentioned issue.

It has often been emphasised that apes do not point declaratively in their natural habitat. This means that they do not point to inform, that is, to the benefit of others. They point imperatively to initiate an action that they themselves in a foreseeable future will benefit from. Experiments have been designed to decide whether apes, even though not pointing declaratively amongst themselves, would understand declarative pointing by humans. It seems that human-raised apes can learn to do so, although the evidence for that is not unequivocal. But why do apes not point to inform? The answer may be that pointing only receives a meaning for apes when it occurs in a competitive context, that is, in a context of action, where the indication receives its meaning from the range of available resources.

Non-competitive or, as we will call it, collaborative co-operation occurs when co-operation does not take the form of an alliance against some other group. The collaboration is not introduced in order to compete for given resources. This means that the utilities in the environment that function as goals are not part of an existing pattern of co-opeartive behaviour in the group or among groups. Then the goal can not receive its value by being an entrenched, limited resource competed for by others. Instead it receives a value independently of the established values in the shared and limited economy of competing agents. Its value will depend on if it can be incorporated among existing values, or if a new value structure will have to be created around it.

Obviously, when there is no established competition for the goal, its value is difficult to determine. It has to be estimated with regard to possible outcomes in the future. Calculating future values demands cognitive resources that apes seem to lack. In support of this, it should be noted that collaborative co-operation has not been observed among apes. This is an important difference between

humans and apes. One reason why apes do not have collaborative co-operation may well be that such co-operation does not make sense to apes, because it does not serve a given end. The end is not prescribed by a competitive context or by the existing behavioural patterns of the group.

Collaborative co-operation is directed at goals that are not among the ones that are competed for in the normal case. It is made possible by the capacity to imagine things that do not occur as part of the normal behaviour. This capacity requires that one can use detached, or context-independent, representations when planning to co-operate. How such representations work will be further described in the next section.

In summary, apes co-operate in order to win something that is available and accessible to them, but not yet in possession. Humans, on the contrary, can as well co-operate in order to achieve something that is so far only desirable and not manifest, and that does not have an entrenched, determinate value. The bases for the latter capacity will be fleshed out in the following sections.

3. Means and ends

As mentioned above, co-operation involves agents, ends (or goals), and means (or strategic actions) to achieve the end. The character of the goal and the ways it can be reached will determine the nature of the co-operation that the agents can engage in. Especially the actuality of the goal is a characteristic of interest for the present discussion.

Goals, say, food, can be characterised either as existing in the present (the ripe apples on the tree in front of you); as not yet existing, but as having precedents, and soon to be realised (the prey that will be hunted and then killed); or as being novel and yet only possible (the fish imagined to live in the unknown river you have just reached). Entertaining the last kind of goal demands that the agent can have goal representations that are detached from the present context, and also from previous contexts.

Detached representations can be increasingly context-free: firstly, they can be independent of the time and place in which their real referent, or what they are about, is situated, and secondly, they can be independent of having a referent at all (G○rdenfors 1996, Brinck and G○rdenfors 1999). In the latter case, they cannot acquire their meaning by standing for some item that exists in the real world. Humans may imagine what is not there, and, moreover, make their imaginations known to each other. Apes, however, do not have detached goal representations (Tomasello 1999). Apes are thus not capable of co-operating in order to reach a common goal that is not present in or implicated the immediate context by (as by learning or memory of previous encounters with similar goals).

The existence of imagined goals introduces a new kind of complexity into cognition. Co-operation builds on social knowledge, that is, knowledge about power relations between agents and about their respective affiliations (Byrne and Whiten 1988). Many situations that involve several agents demand tacit cost-benefit analyses of predicted actions and outcomes before individual action can be taken. Such analyses require that there is information to be had about the agents involved, concerning ranking, kinship, and such. The analyses also require that the agents have knowledge about means and ends.

But if the goal is imaginary, its value is also imaginary. It does not constitute a stable resource and its value cannot be calculated. This has consequences for co-operation. Much co-operative behaviour is compensatory – it is performed in exchange for other actions that have already been performed or are anticipated. Grooming may be performed for the exchange of grooming itself or with the further end of getting support in future fights or alliances. During co-operative hunting, apes help each other because they have a common goal of sharing food.

The exchange that underlies compensation presupposes that the resources are known, or at least can be known. But in the case of imaginary goals, compensation becomes much more of a venture than a safe strategy. If individual agents cannot base their actions on an analysis that takes into account the estimated value of the goal, the collective outcome of strategic actions behind co-operation is shaken.

Similarly to goals, means can be characterised in different ways. While we characterised goals in terms of actuality, we describe means in terms of how they relate to their goal. Means can either be invariably connected to a certain kind of goal; they can be variable, in case there are different ways of reaching the same goal; or they can be novel and original. Tomasello and Call (1997) underline that in understanding behaviour as intentional, one understands

that "different means may be directed toward the same end and that the same means may be used for different ends" (p. 361). As they see it, intentionality involves more than understanding that other agents can generate their behaviour spontaneously and that their behaviour is directed (p. 203).

What is intentional about the capacity to separate means and ends? Intentionality, or goal-directedness, emerges as an ability to adopt the means to the situation at hand. To be reached, the same goal may require different means in different contexts. And the same means may afford different goals in different contexts. Intentional behaviour is flexible. It depends on the agent's ability to appreciate the distinctive character of each context it encounters. It also depends on the agent's ability to learn about new contexts and how to deal with them.

According to Tomasello and Call (1997), chimpanzees cannot represent the goal without the means to reach it. Goal and means form an indivisible whole. Humans, on the other hand, can reflect about different ways of reaching the goal. Since humans also can produce goals that have not yet analogues in the environment, human co-operation has the potential of being very flexible.

4. Co-operation and communication via joint attention

Co-operation depends on communication. Agents will not be able to co-ordinate their actions, unless they can communicate about their present state and the action that will follow upon it. They need to communicate about the goal of the co-operation, as well as about their own and each other's means to reach the goal. In many cases, communication can rely on display and signalling of the present and the next behavioural states. But sometimes co-operation stands in need of anticipation. This is the case when the goal is not present in, nor implicated by the actual context. It also happens when there are several possible goals, but none actual. Then communication cannot simply rely on contextual means to transmit information between agents.

Signalling depends on causal influence between agents. Information can be transmitted causally by perceptual cues. The agents communicate by signalling their own states and action-readiness to each other. Co-operation sometimes does not need more than a chain of signals to issue in quite complex behaviour. The actions directly follow upon each other because they causally prompt each other. For example, a flock of flying geese in a V-formation seem to be involved in fine-tuned co-operation, but the formation is actually caused by the aerodynamic forces which make the V-formation the least strenuous for the geese.

But sometimes co-operation is more elaborated. For instance, perceptual cues of a certain type may prompt several kinds of behaviour. The appropriate behaviour should be chosen with regard to the situation at hand and the goal. In order to produce the appropriate behaviour, the agents will have to check their mutual reactions to each other and to the goal. Consequently, for the co-operation to be successful, the agents must engage in joint attention. It is triadic: agents attend not only to a shared object, but also to the attention of each other. Joint attention allows for two or more subjects to focus their perception simultaneously on a single attentional object (so-called mutual object-focused attention) provided that the subjects have focused on each other beforehand (so-called subject-subject attention) (Bruner 1998, Tomasello 1999, Brinck 2001).

In mere subject-focused attention, attention is directed at the behaviour of other subjects. By, for instance, looking at each other, two subjects can detect their respective attentional objects. They do so on the basis of the direction of their respective movements in combination with a salient object that functions as a target. Co-ordination of attention based on saliency and behavioural co-ordination result in mutual object-focusing. Mutual object-focusing can spread automatically among subjects.

Joint focusing of attention is, as mentioned above, based on subject-subject attention and requires agents who actively seek goal-oriented information. It involves attentional co-ordination in the sense that the subjects attend to each other as subjects capable of attending. One might think that this necessarily involves having a theory of mind and being able to grasp that other subjects have mental states, but this is not the case. An agent can read the attention of the other agents from their gaze and behavioural cues. The awareness of states of attention relies on evidence such as facial expression, gaze, and body posture that is related to a certain kind of action readiness and vigilance. Joint attention rests both on information laid out in the environment and on such attributed to mental states. Awareness of the capacity to attend in others consists

in having an expectation of others to manifest attentional behaviour.

Agents capable of subject-subject attention primarily attend to each other's attentional states, not to the behaviour that is a consequence of attention. However, they do not have to extract information about the attentional states of others explicitly as being about such states (which may have intentional components), but can do so by monitoring such information as it is manifested in intentional, bodily behaviour. On the basis of such information, they can engage in attention contact, during which they simultaneously check each other's state of attention, for instance, by eye contact (GÛmez 1994).

Agents that can engage in subject-subject attention also attend to each other as capable of attending in a goal-intended way, that is, in a way that is not controlled by the object of attention. Goal-intention provides for the capacity to either direct or follow the attention of the other subject in the absence of salient objects (though not in the absence of objects altogether). This means that attention focusing can be guided by the subjects' mutual attention to each other instead of by what is happening around them.

Joint attention may occur in both competitive and collaborative co-operation. This capacity makes possible ways of co-operating that cannot occur without it. An example would be, when hunting a herd of antelopes, two agents can jointly attend to the same individual in the herd and thereby co-operate by taking different roles in killing the selected antelope. It has been argued that apes are not capable of joint attention (Tomasello 1999). If this is true, they would not be able to take part in this more advanced form of hunting.

5. Co-operation and communication by symbols

If the desired goal is not present in the actual context of the agents, the process of attention reading will not work. In order to communicate about absent goals, the agents must first be able to entertain detached representations, and then be able to attend jointly to such representations. It seems hard to explain how this is done without evoking symbolic communication.

Tomasello (1998) describes symbols as social, inter-subjective, and bi-directional. Bi-directionality assures that a competent user who understands somebody else's use of a symbol as having a particular content can herself produce the symbol with the same content.

Production and comprehension of symbols go together. Deacon (1996, 1997) explains how symbols acquire meaning by being related to each other in various ways, like opposition, substitutability, and adjacency. Symbolic reference depends on indirect reference to objects, with the help of other symbols.

Communication by symbols is quite intricate, because the meanings of the symbols are general and defined by interrelation. It has so far not been shown that apes can communicate in a fully symbolic way (Deacon 1997, Tomasello 1999). On the contrary, it seems that apes in their natural habitat, which have not been trained, mainly exploit indexicals in sign communication. That means that the signs the apes use acquire their meaning by standing in a causal and contingent relation to what the signs are about. Indexicals are dependent on the context for their meaning. This has the side effect that the same sign, or gesture, may be used with different meanings in different contexts. One can raise the arm to indicate, for instance, both danger and food. This stands in contrast to symbols, which retain their meaning across contexts.

The use of symbols in thought has the consequence of setting the agent free from desires imposed by the actual context. For instance, it may be hard to give up a utility in possession for a future, but more precious one. The reason is that the present utility seems so much more valuable than the one that can only be imagined. It has been shown that apes who have been trained to use symbols to communicate, instead of the things themselves, are able to plan for future desires, and not to give in for present ones.

This is illustrated by an experiment with chimpanzees performed by Boysen and Bernston (1995). They put peanuts in two heaps of different size placed on a table out of reach of the apes. One ape was to point at one of the heaps, and then that heap was given to the other ape, while he himself got the one that he did not point at. The result of the test was surprising. The ape repeatedly pointed at the biggest pile and was very disappointed when that pile was given to the other one, and he himself received the small pile.

The chimpanzees consistently chose the bigger pile of food. The presence of the desired food seem to make them incapable of imagining the near future when the other party receives the pile that they choose and they are left with the other pile. Boysen and Bernston's experiment clearly shows how difficult it is to manage even the simplest form of planning for a future goal. Deacon (1997,

in having an expectation of others to manifest attentional behaviour.

Agents capable of subject-subject attention primarily attend to each other's attentional states, not to the behaviour that is a consequence of attention. However, they do not have to extract information about the attentional states of others explicitly as being about such states (which may have intentional components), but can do so by monitoring such information as it is manifested in intentional, bodily behaviour. On the basis of such information, they can engage in attention contact, during which they simultaneously check each other's state of attention, for instance, by eye contact (GÛmez 1994).

Agents that can engage in subject-subject attention also attend to each other as capable of attending in a goal-intended way, that is, in a way that is not controlled by the object of attention. Goal-intention provides for the capacity to either direct or follow the attention of the other subject in the absence of salient objects (though not in the absence of objects altogether). This means that attention focusing can be guided by the subjects' mutual attention to each other instead of by what is happening around them.

Joint attention may occur in both competitive and collaborative co-operation. This capacity makes possible ways of co-operating that cannot occur without it. An example would be, when hunting a herd of antelopes, two agents can jointly attend to the same individual in the herd and thereby co-operate by taking different roles in killing the selected antelope. It has been argued that apes are not capable of joint attention (Tomasello 1999). If this is true, they would not be able to take part in this more advanced form of hunting.

5. *Co-operation and communication by symbols*

If the desired goal is not present in the actual context of the agents, the process of attention reading will not work. In order to communicate about absent goals, the agents must first be able to entertain detached representations, and then be able to attend jointly to such representations. It seems hard to explain how this is done without evoking symbolic communication.

Tomasello (1998) describes symbols as social, inter-subjective, and bi-directional. Bi-directionality assures that a competent user who understands somebody else's use of a symbol as having a particular content can herself produce the symbol with the same content.

Production and comprehension of symbols go together. Deacon (1996, 1997) explains how symbols acquire meaning by being related to each other in various ways, like opposition, substitutability, and adjacency. Symbolic reference depends on indirect reference to objects, with the help of other symbols.

Communication by symbols is quite intricate, because the meanings of the symbols are general and defined by interrelation. It has so far not been shown that apes can communicate in a fully symbolic way (Deacon 1997, Tomasello 1999). On the contrary, it seems that apes in their natural habitat, which have not been trained, mainly exploit indexicals in sign communication. That means that the signs the apes use acquire their meaning by standing in a causal and contingent relation to what the signs are about. Indexicals are dependent on the context for their meaning. This has the side effect that the same sign, or gesture, may be used with different meanings in different contexts. One can raise the arm to indicate, for instance, both danger and food. This stands in contrast to symbols, which retain their meaning across contexts.

The use of symbols in thought has the consequence of setting the agent free from desires imposed by the actual context. For instance, it may be hard to give up a utility in possession for a future, but more precious one. The reason is that the present utility seems so much more valuable than the one that can only be imagined. It has been shown that apes who have been trained to use symbols to communicate, instead of the things themselves, are able to plan for future desires, and not to give in for present ones.

This is illustrated by an experiment with chimpanzees performed by Boysen and Bernston (1995). They put peanuts in two heaps of different size placed on a table out of reach of the apes. One ape was to point at one of the heaps, and then that heap was given to the other ape, while he himself got the one that he did not point at. The result of the test was surprising. The ape repeatedly pointed at the biggest pile and was very disappointed when that pile was given to the other one, and he himself received the small pile.

The chimpanzees consistently chose the bigger pile of food. The presence of the desired food seem to make them incapable of imagining the near future when the other party receives the pile that they choose and they are left with the other pile. Boysen and Bernston's experiment clearly shows how difficult it is to manage even the simplest form of planning for a future goal. Deacon (1997,

p. 414) writes that the choice is difficult for chimpanzees since the indirect solution (choosing the small pile) is overshadowed by the direct presence of a more attractive stimulus, namely, the big pile. They cannot suppress their perception.

If one performs the same kind of experiment with human children, they have no problem choosing the small pile – from the age of two and up. They can imagine receiving the big pile when they point at the small one. When children are younger they behave more like chimpanzees. It has been shown that persons with autism have difficulties that closely resemble those of chimpanzees and small children (Russell *et al.* 1991).

The chimpanzees in Boysen's study were also trained to connect numerals with quantities, so that they knew that the figure 5 corresponded to a bigger pile of sweets than the figure 3. Once they had learned the meaning of the numerals, they were instead allowed to choose between two cards showing numerals, and the pile corresponding to the selected numeral was given to the other chimpanzee, while they themselves received the pile corresponding to the numeral they had not chosen. The results were now better: they could learn to select cards so that they received the big pile. The explanation is that there is no longer a conflict between the symbols they see and what they have to choose – the cards had no intrinsic value for the chimpanzees. This variant of the experiment is a form of conditioning, which is no problem for chimpanzees.

Why is it cognitively more difficult to plan for future needs than for current ones? The answer has to do with the different representations that are required for the two types of planning. When planning in order to satisfy current needs, one must be able to represent actions and their consequences, and to determine the value of the consequences in relation to the needs one has at that moment. But no separate representation of that need is required. To plan for future needs, on the other hand, one must also be able to represent these potential needs (and to understand that some of them will arise). The available ethological evidence so far indicates that man is the only species of animal with the ability to imagine future wishes and to plan and act accordingly (Gulz 1991). Deacon (1997) calls our thinking front-heavy; anticipatory planning takes place in the frontal lobe, the most recently evolved part of the brain.

Human language is the prototype example of a symbolic communication system. Clearly, human language paves the way for

long-term co-operation and for co-operation towards as yet unrealised, perhaps indeterminate, future goals. The goals and the means to reach them, as well as the agents' expectations and misgivings concerning the co-operation, can be externally shared through the linguistic medium. This kind of sharing gives humans an enormous advantage concerning co-operation in comparison to other species.

Humans can co-operate even when the desirable goal is remote from the conditions set by the environment. Since humans can have detached representations of both means and goals, and also can separate means from goal representations, the possibilities that co-operation opens up for the future are extensive. The environmental conditions can be completely transformed in order to give way for the means considered necessary to produce or realise an imaginary goal.

Humans have often used such tactics in history. Whole landscapes have been not only transformed, but even created from scratch to make room for new inventions and the realisation of new desires. One example is the capital of Brazil, Brasilia, that was built in the end of the 1950s by Costa and Niemeyer. The design of the city was guided by a vision of a modern and technologically based future. Another example is the artificial irrigation that has made possible agriculture in desert landscapes, and yet another the construction of water pools or reservoirs to either exploit or change the reaches of rivers.

REFERENCES

Boysen, S. and G. Bernston, 1995. "Responses to quantity: perceptual versus cognitive mechanisms in chimpanzees (Pan troglodytes)", *Journal of Experimental Psychology and Animal Behaviour Processes* 21 pp. 82-86.

Brinck, I. 2001. "Attention and the evolution of intentional communication", *Pragmatics & Cognition* (to appear).

Brinck, I. and P. Gärdenfors 1999. "Representation and self-awareness in intentional agents", *Synthese 118*, pp. 89-104.

Bruner, J. 1998. "Routes to reference", *Pragmatics & Cognition* 6, pp. 209-227.

Byrne, R. W. and A. Whiten, eds. 1988. *Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes, and Humans*, Oxford University Press, Oxford.

Deacon, T. 1996. "Prefrontal cortex and symbol learning" in *Communicating Meaning: The Evolution and Development of Language.* eds. B. M. Velichkovsky and D. M.

Rumbaugh, Lawrence Erlbaum: Mahwah, NJ.
Deacon, T. 1997. *The Symbolic Species: The Coevolution of Language and Brain.* MIT Press: Cambridge, MA.
Gärdenfors, P. 1996. "Cued and detached representations in animal cognition", *Behavioural Processes 36*, pp. 263-273..
Gómez, J. C. 1994. "Mutual awareness in primate communication: A gricean approach", in *Self-Awareness in Animals and Humans.* eds. S. T. Parker, R. Mitchell, and M. Boccia, Cambridge University Press: Cambridge.
—— 1998. "Some thoughts about the evolution of LADS, with special reference to TOM and SAM", in *Language and Thought.* eds. Carruthers, P. and J. Boucher, Cambridge University Press: Cambridge.
Gómez, J. C., E. Sarriá, and J. Tamarit 1993. "The comparative study of early communication and theories of mind: Ontogeny, phylogeny, and pathology" in *Understanding Other Minds. Perspectives from Autism.* eds. S. Baron-Cohen, H. Tager-Flusberg, and D. J. Cohen, Oxford University Press: Oxford.
Gulz, A. 1991. *The Planning of Action as a Cognitive and Biological Phenomenon*, Lund University Cognitive Studies 2, Lund.
Russell, J., N. Mauthner, S. Sharpe, and T. Tidswell, 1991. "The 'windows task' as a measure of strategic deception in preschoolers and autistic subjects", *British Journal of Developmental Psychology 9*, pp. 331-350.
Tomasello, M. 1996. "Do apes ape?", in *Social Learning in Animals. The Roots of Culture.* eds. J. Galef, and C. Heyes, Academic Press: New York, NY.
—— 1998. "Reference: Intending that others jointly attend" *Pragmatics & Cognition* 6, pp. 229-243.
—— 1999. *The Cultural Origins of Human Cognition*, Harvard University Press, Cambridge, MA.
Tomasello, M. and J. Call, 1997. *Primate Cognition.* Oxford University Press: New York, NY.
de Waal, F. B. M. 1982. *Chimpanzee Politics.* Jonathan Cape: London.

Language and Thought

Language and Thought
A Generative Linguistic Perspective

B.N. PATNAIK

Since consciousness is the focal theme of the seminar, I should like to start with the observation that to the best of my knowledge there is no theory of consciousness that incorporates a theory of language in any illuminating manner, and that there is no linguistic theory that relates in any insightful way to a theory of consciousness, if such a theory exists, using the word "theory" in the sense in which it is used in the natural sciences. However a history of thought would show that there have been attempts to link language, by which is meant the articulated expressions, thought, the pre-verbal stage, and consciousness, the pre-thought stage, in a manner of speaking. Bhartṛhari's is one such attempt. Now, given that for hundreds of years it has been held that language is a vehicle for the expression of thought – "Language is a tool for thought", as Chomsky would say – it might appear to be rather curious that modern generative linguistics has nothing really substantial to say by way of explicating the relation between language and thought. By "modern generative linguistics" is meant here the so-called "Chomskyan" linguistics because it has always been the mainstream generative linguistics ever since the beginning of modern generative linguistics in the early nineteen fifties. It is also a fact that the other schools of generative linguistics (practitioners of, say, Lexical Functional Grammar, Generalized Phrase Structure Grammar, or Case Grammar, etc.) have nothing or very little indeed to say when it comes to the meta-theoretical issues pertaining to generative linguistics. This fact that generative linguistics offers no or virtually no insights into the relation between language and thought needs an explanation, which this paper attempts to provide. Needless to say, the explanation offered here is just one explanation. This is the main aim of the paper. The other aim of the paper is to examine whether Bhartṛhari's idea mentioned above can be translated to some extent at least into a modern conception regarding the link

that connects the pre-thought, thought and language stages in some insightful way. The paper outlines the broad framework of generative linguistics first and then goes on to articulate an explanation.

It would perhaps be in the interests of the presentation if we present Bhartṛhari's ideas regarding language, thought and consciousness, as we put it, before we address ourselves to the central concern of this paper. This great ancient Indian linguist and philosopher of language had roughly the following to say regarding the process of production of language from the speaker's point of view: some unstructured and undifferentiated "feeling" or "experience" takes the form of structured and differentiated thought, which then gets transformed into language. Matilal (1990, p.88) puts this idea thus: ". . . Bhartṛhari has posited three stages of language or speech. The first stage, where there is complete identity of language and thought is called the *pasyanti* stage; we can call it 'non-verbal'. The 'intermediate stage', where despite the identity of thought and language their difference is discernible, can be called the 'pre-verbal' stage. And the third, the *vaikari* stage, can be called the 'verbal' stage." The following is somewhat more informative about the so-called 'intermediate stage': "Before the proper articulation of the sound-sequence or utterance, there is another 'intermediate' stage (called *madhyamā vāk*) where the language and the thought it conveys are still one and undifferentiated, but at this 'pre-verbal' stage the speaker sees them as differentiable. In other words, he recognizes the verbal part, which he is able to verbalize either to himself or to another, as distinct and separable from the *artha*, 'meaning' or 'thought' (ibid. 86)". Our purpose in bringing Bhartrhari into the ambit of discussion here is to draw attention to the fact that the relation between language and thought is an ancient concern, at least fifteen hundred years old. Down the centuries, no one has held the view that there is no relation between the two, and at the same time no one would assert that it is a question to which a satisfactory answer has been found. To us it appears that post-Bhartṛhari scholarship has not really increased our understanding of this relation in any significant way – Fodorians might differ, but we will not dwell on this matter here in any detail. Several questions readily arise when one tries to understand the relation between language and thought: what, for example, are the categories of thought as distinct from those of language? What is the internal structure of a unit of thought? Is the notion of "complete thought" non-trivially different from that of "sentence"?

For the purpose of drawing a broad outline of modern generative linguistics, relevant for the present paper, we choose the version of generative grammar known as the "Minimalist Programme" (MP), as it was articulated in Chomsky (1995 (b)) in particular. The basic question that generative linguistics – sometimes called "bio-linguistics" – addresses itself to is the following: what is the innate, biological endowment that humans, and only humans, have that makes them acquire the knowledge of a language and put this knowledge to use? (There is a subsequent question to this, namely, what is the physical basis of this knowledge? But this is indeed an "integration" question – integration between the present-day linguistics and the present-day biology, or rather the brain sciences – one that is conceivable today as a legitimate question, although it is quite unclear as to what form the answer would take.) It is worth mentioning that whereas contemporary generative linguistics offers valuable insights as to how humans acquire the knowledge of their languages, their mother tongues, it offers almost nothing insightful regarding how they put this knowledge to use in negotiating with the world. The biological endowment which linguists believe to be species specific is called "the initial state of knowledge of language" ("universal grammar" in another terminology), and the knowledge of their own language that they acquire as they are exposed to a specific linguistic environment is sometimes called "the steady state" knowledge ("core grammar" of their language or "I-language" in other terminologies). There are various intermediate stages, as the initial state of the knowledge of language in a human develops into the steady state of a particular language. The generative linguistic enterprise prioritizes research on the initial state, but since theoretical proposals concerning the initial state or universal grammar can be evaluated on the basis of I-language facts (and related) facts alone, theories of I-languages inevitably receive significance. Now one basic aspect of the question regarding the nature of the initial state or the universal grammar is the following: what is the architecture of this state (equivalently, what is the internal structure of universal grammar)? (The other is: how is this design, this architecture, activated? The answer to this question at one level seems to be rather straightforward, namely, as a result of exposure to language. We do not propose to say anything more about it here.)

Chomskyan linguistics subscribes to a modular theory of the mind,

and maintains that mind has a language module, a "language faculty", a "mental organ", in another terminology. The language faculty is a distinct faculty because it includes principles that are specific to it. For instance, the visual faculty does not include binding principles or the principle of structure dependency, which are part of the language faculty, but it includes the rigidity principle and those governing apparent motion, not available to the language faculty (see Chomsky 1988: 159-161)). However, although a separate module, the language faculty interacts with other systems of the mind: "external systems", in the Minimalist terminology. The language faculty generates formal objects, which are pairs of expressions, and these expressions receive interpretation in these external systems: the articulatory-perceptual (AP) system and conceptual-intentional (CI) system. The former assigns a phonological interpretation to the input expression, and the latter, a semantic interpretation to the input expression. The so-called CI component can be thought of as the component of thought in the more familiar terminology. This is the way the relation of thought and language is conceptualized in the MP model. It may be noted that what is specific to the human species is the language faculty, which means that some other species might have the capacity for thinking and feeling, but they do not have the capacity, the equipment, to articulate the same in terms of language, which has the basic properties of discrete infinity and displaced reference. "Displaced reference", a typical feature of a natural language, refers to the property that an entity, which is pronounced at one place, may not be interpreted at that place. For instance, in "What did you see?" the word "what" is pronounced in the beginning of the sentence but is understood as the object of "see". Since the object of the verb immediately follows the verb in English, it is evident that the object has been dislocated from its place to the beginning of the sentence. If humans interacted in telepathy, then the AP component would be unnecessary; there will be no need to assign phonological interpretation to the expressions generated by the language faculty.

The language faculty has a lexicon and a computational component. The lexicon is a repository of lexical items, each a bundle, an assemblage, of phonological, syntactic and semantic features, and is dedicated to the language faculty. A computation, called "select", selects lexical items from the lexicon, and deposits them in, say, a lexical space. Then a computation called "merge" systematically constructs larger units, phrases, in a permissible way, determined by the properties of the entities involved, with the result that the internal structure of a phrase, which is a projection of a head, (for example, the verb phrase, VP, is the maximal projection of the verb, V, its head) is manifest. This operation must exhaust all the lexical entities selected from the lexicon, for the representation to converge. Then the structure, the "initial phrase marker" in some sense, is available to the computational component that displaces features (consequently, lexical items), checks features, erases certain specific ones, etc. The various computations at this stage are the consequences of the functional features (such as Case features—nominative, oblique, etc.—of nouns) of the lexical items involved. As far as the derivation for semantic interpretation is concerned, when the computation is complete, the resultant structure, "the finally derived structure" ("the interface level representation") is fed into the external system of thought (the CI system) for interpretation. This outline of the architecture of the language faculty is adequate for the purposes of the present discussion; for the same reason, we will say nothing more about the derivational mechanism. We now turn to the CI interface representation.

Of direct interest for us is the following requirement of the interface representation: it must be legible to the external system of thought so that it can receive an interpretation. Very little is known about the external system, which is to be expected, because there is as yet no theory of the CI external system, this knowledge system that uses knowledges other than the knowledge of language (for just one example, the world knowledge), to assign interpretation to the input representation. If little is known about the external system of thought, what sense, one might ask, does it make to impose the legibility condition mentioned above? There is certainly justification for such a requirement since the output of the language faculty has to be used; the knowledge of language is not a piece of decoration in the mind! Chomsky seeks to resolve this problem in the form of a methodological imperative: the quest into the nature of the language faculty cannot be delayed because very little is known about other relevant faculties. One might work under the assumption that the LF representation, the interface representation under discussion, is somehow legible to the external system of thought. Does it mean that the units and the structure of thought are isomorphic with the same of language? Observes Neil Smith, "Öit looks plausible to claim

that the structure of sentences is overwhelmingly similar to the structure of thoughts, taking it for granted that thoughts *have* structures. (1999, p. 47)"

But this is intuitively unsatisfactory. Compare, for instance, the notions of sentence and proposition, which for probably centuries were viewed as essentially the same notion, which may account for the traditional conception of a sentence as a unit that expresses complete sense. Consider the sentence "Manisha Koirala is the thinking man's Madhuri Dixit". Is it a single proposition or does it have more than one proposition? This sentence telescopes an assertion about Manisha Koirala, one about Madhuri Dixit, and a comparative assertion by specific groups of people — at least these. Whatever be the propositions, the sentence clearly contains more than one proposition. (Incidentally, some schools of generative linguistics, among them generative semantics school of the sixties, have posited derivations that reflect the idea that underlyingly a sentence is a single or a complex set of propositions. Had it succeeded, the approach that generative semantics followed would have led to the result that the logical and the grammatical categories are indeed the same at the deepest level of abstraction. It is only at the surface level that they are different. But generative semantics was an attempt that failed. The mainstream generative linguistics today would not derive this sentence from a multi-clausal (or sentential, without differentiating between "sentence" and "clause" for theoretical purposes) base. Similarly, is "chase" an atomistic concept as "climb"? It is not, because, as we know, "chase" comprises the concepts of "follow" and "intention" (see Chomsky 1988: 32-34).

Consider the relation between concept and word from another point of view. It has been observed that very young children, one-and-a-half-year-olds, have a considerable size of vocabulary. This is surprising, because it cannot be a result of instruction. So like the acquisition of syntax, acquisition of vocabulary is possible because of the biological endowment. Now what the child has by virtue of this endowment cannot be words, because words belong to particular languages: "climb" belongs to English, and *cadnā* to Hindi. One could therefore surmise that the human is born with a repertoire of concepts. So what she has to learn from the linguistic environment is the linguistic form in which a certain concept is realized in a certain language– say, the concept CLIMB is realized in English as the word "climb", and in Hindi as the word *cadnā* This means that the word-

concept (or language- thought) relation is as follows: if there is a certain word in a language, it is because there is a corresponding concept for it in the repertoire of concepts. Needless to say, from this it does not follow that a concept, which is part of our biological endowment, has a lexical realization in every language. There is however a problem in this way of viewing the relation between lexical items and concepts. We surmise that there is a certain concept because there is a lexical item – it is inconsequential that this lexical item may occur in the lexicon of just one language – in other words, we have no word- independent way of knowing whether there is a certain concept in the repertoire of concepts or not. We know there is a certain word "climb" because there is a concept called CLIMB, and we know there is a concept called CLIMB because otherwise there wouldn't have been the word "climb" in English (see Chomsky (1988), (1997) and (2000), among others). It must be noted that it does not follow from the above that there is no mental repertoire of concepts. What it shows is that we do not know what precisely these concepts are, what their internal structures are, and how these concepts are organized in the repertoire, how atomistic concepts are combined to form complex ones (presuming from the discussion of "chase" and "follow" above that indeed concepts can be atomistic and complex), how a concept relates to a word and a great deal else. In short we do not, as yet, to the best of our knowledge, an illuminating theory of thought.

We might mention here in passing that it should not give rise to any skepticism regarding the point we were making in the discussion above regarding words and concepts that we did not deal with words referring to concrete objects. No semantic theory directly relates words and the "objects" they refer to in the real world; those like semiotics or ancient Indian theories of meaning, which aim to understand this relation do so in terms of word-concept-object relation. Thus our not bringing into the discussion words and concrete objects do not weaken our point in any way. In any case the relation between words and real world objects is much more complex than is often believed to be, as Chomsky points out (Chomsky 1993). To give just one illustration, an artifact having a certain set of specifications can be called "table" or "bed" depending on how it is put to use: if a dwarf uses it for the purpose of sleeping, she would call it a "bed", whereas an ordinary person, who would use it for normal work would call it a "table".

The lack of a theory of thought must not impede progress in linguistics, as Chomsky observes, therefore one explores the nature of the language faculty assuming that there is a fit between the representation at the CI interface, sometimes called the LF representation, and the readability requirements of the thought component (the CI component). One assumes further that an explication of this fit is a solvable problem, which would achieve resolution with the increase of our understanding of the nature and the structure of thought. It is only when we have a much clearer understanding of this fit that we can hope to achieve some significant understanding of the relation between language and thought. This, to our mind, is in essence, the position of contemporary generative linguistics on the relation between language and thought.

We now return to Bhartṛhari. We know that he posited an antecedent state to what for him would be the state of structured thought, which we have called "thought" in the discussion on generative linguistics: as we know, this antecedent state is the *pasyanti* state and the state of structured thought is the *madhyamā* state. (He posited yet another state, *parā*, which is the antecedent state to *pasyanti*. We will say nothing about *parā* because it is inconsequential for our present purpose. If the *pasyanti* state is accessible to the great seers by *savikalpa samādhi*, the *parā* state, the state of realization of *sabdabrahman* is attained by a seer in the state of *nirvikalpa samādhi*, which can be said to mean that it is beyond ordinary human grasp (see H.G. Coward and K. Kunjunni Raja, p. 328).) The following provides us with one characterization of *pasyanti*: "*Pasyanti* is the direct experience . . . of meaning as a noumenal whole. At this level there is no distinction between the word and the meaning, and there is no temporal sequence . . . *pasyanti* is, by definition, beyond the level of differentiated cognition. . . . It occurs at the level of direct intuition and therefore must finally be understood through experience (ibid, p.61)." This state captures Bhartṛhari's idea of where pre-linguistic or pre-verbalization thought comes from. Incidentally, that there is thought prior to verbalization would not be unacceptable to the present-day generative linguists, although a model of grammar, the MP model for example, does not formalize this idea in the form of a thought component generating thoughts subsequently verbalized in say, a language component. The reason is that this is not the insight that is at the core of contemporary thinking on grammatical theory. It is, as we know, something else,

namely that there is a language faculty, which constitutes a separate knowledge-module, and the output of this module is what is accessible to the thought component, where is receives interpretation. And dealing with the pre-verbalization state indeed is something that does not come within the domain of generative linguistics.

Stripping away all the metaphysical associations and significance from the notion of *pasyanti*, we could view it as a state of raw feeling, a kind of purely affective experiential state perhaps, a state where the experiencer and the experience are not separate. It might be argued that this would be an inexact characterization of this notion since what we have stripped away is an integral part of it. This objection may not be without substance; after all, particular theoretical notions belong to specific theories and may lose their descriptive or explanatory value outside those theories. But at the same time one does often try to seek similarities between theoretical constructs, make connections in the process, in order to achieve a perhaps more general understanding of things. Taking the risk of saying the obvious, thereby inviting censor, we might observe that this comparative perspectivet – his attempt to connect, to seek similarities and parallels – has really enriched us. Theta theory, to give an example, belongs to grammatical theories constructed within the so-called "Principles and Parameters" (PP) approach, and the *kāraka* theory belongs to a different grammatical system, namely, the ancient Indian, and surely *kartā* cannot be understood as the same notion as either subject or agent, ("agent"as the interpreted notion) of the PP grammars. But trying to see the *kaaraka* theory from an understanding of the theta theory or the converse might provide us with an understanding of how different grammatical theories have tried to execute the same deep, or "leading", in a different terminology, idea or insight. Although *ākānsā* ("expectancy") – roughly speaking, the notion that words in a sentence have mutual expectancy (the verb *pāc* "cook", for instance, expects a *kartā* and a *karma* in a well formed sentence) – and "s-selection" (informally speaking, the idea that lexical items (predicates) have theta grids associated with them ("kill" must have an agent and a theme in a sentence for its meaning to be saturated) —are arguably not identical notions (for reasons we cannot and need not go into here, since these involve details), but these are clearly similar, different executions of the same (or very nearly the

same) leading idea. If we refuse to see these from a comparative perspective on the ground that they belong to two different theories rooted again in two different intellectual milieus, we will not be able to raise the question as to why different grammatical theories need similar theoretical notions, as shown above. Such examples can be readily multiplied. In sum, the above shows that trying to see Bhartrhari's notion of *pasyanti* without its metaphysical associations in order to draw some parallel between his way of thinking and the modern one on a similar issue does not deserve to be censored outright. So for our present purposes it is an experiential state where the experiencer and the experience are indistinguishable — a state of raw feeling, a state of pre-thought/cognition consciousness.

There is then the move from this state to the *madhyamaa* (intermediate) state, and this is where the experience comes under the cognitive gaze, as it were; experience becomes dissociated from the experiencer, and thought gets fragmented into categories. One articulation of the reason why this happens is as follows: "... there is present at this level (*pasyanti*, clarification, mine) a kind of "going out" or desire for expression. This impulse is the *pratibhā* "instinct", which in one sense may be said to motivate the phenomenalization into sentences and words of the *pasyanti* vision, so that communication may occur (ibid, p.61)." We might understand a state of consciousness undergoing cognitive scanning because of the way the human mind works. We may not, however, subscribe to the idea of communication intention as the sole cause for the pre-thought state to undergo the stage of thought because there seems to be no compelling reason to do so. Besides there is a methodological problem for the generative linguist as far as subscribing to the role of intention is concerned. A generative linguist is committed to the approach that the framework for the study of what lies "above the neck" and what lies below it must be the same. Within this approach there is no place for intention. The perspective of the natural scientific approach is that a system functions because of its own internal mechanisms, and not because of the triggering role of intention (for a succinct statement about the generative linguist's method for the study of mind, one component of the mind-brain really, namely, the language faculty, see Chomsky (1995 a).

Bhartṛhari articulates no theory of thought, as far as we can see. There is no theory of thought that he presents that raises and tries to answer the questions regarding concepts enumerated earlier here.

Notions such as time, action, state, and relations such as *kartā*, *karma*, etc. of ancient Indian linguistics—entities and relations such as these are what are phonologically realized—are notions of grammar, not of thought; they would belong to the theory of the CI interface in terms of the MP framework (oversimplifying the matter a bit perhaps). Our understanding of the CI interface is too limited and ad hoc, since, it may be recalled, there is as yet no clear understanding of the nature and the mechanism of the perfect fit between the requirements of the external system of thought and the output expressions of the computations internal to the language faculty, which, as already mentioned, are inputted into the external system of thought.

In sum, what we have mainly considered here is how far one can capture, with the theoretical resources available within the natural scientific approach followed by modern generative linguistics, the centuries-old view about the relation between language and thought, namely, that language is an expression of thoughts and feelings. We have noted that thought and language are intuitively related, but at the present stage, our understanding of the nature and structure of thought is so very insignificant that one is hardly in a position to make any interesting theoretical observations about this relation.

Sincerest thanks are due to Bijoy Boruah, Nirmalangshu Mukherji, Shobini Rao, and Achla M.Raina for some helpful discussion on certain ideas discussed here.

REFERENCES

Chomsky, Noam (1972). *Language and Mind*. New York: Harcourt Brace Jovanovich. (expanded edition)

—— (1986). *Knowledge of Language*. New York: Praeger.

—— (1988). *Language and Problems of Knowledge*. Cambridge: M.I.T. Press.

—— (1993). *Language and Thought*. London: Moyer Bell.

—— (1995a). "Language and Nature". Mind, January, pp. 1-61.

—— (1995b). *The Minimalist Program*. Cambridge: M.I.T. Press.

—— (1997). "Language and Mind: Current Thoughts on Ancient Problems". (Mimeograph)

—— (2000). (Mukherji, N., B.N.Patnaik, R.Agnihotri eds.) *The Architecture of Language*. Delhi: Oxford University Press.

—— (2001a). "Beyond Explanatory Adequacy". (Mimeograph)

—— (2001b). "Language and Rest of the World". Lecture delivered in Delhi on November 4. (cassette)

Coward, Harold G. and K.Kunjunni Raja (eds.).(2001) *The Philosophy of the Grammarians.* (reprint) (Volume 5 of *Encyclopedia of Indian Philosophers*). Delhi: Motilal Banarsidass.
Fodor, J. 1975). *The Language of Thought.* New York: Crowell.
Jenkins, Lyle (2000). *Biolinguistics.* Cambridge: Cambridge University Press.
Matilal, Bimal K. (1990). *Word and the World.* New Delhi: Oxford.
Smith, Neil (1999). *Chomsky: Ideas and Ideals.* Cambridge: Cambridge University Press.

Recognition

SUPRIYA CHAUDHURI

In an essay on the uncanny, written in 1919, Freud records a disquieting moment of personal experience in the course of a railway journey:

> I was sitting alone in my *wagon-lit* compartment when a more than usually violent jerk of the train swung back the door of the adjoining washing-cabinet, and an elderly gentleman in a dressing-gown and a travelling cap came in. I assumed that he had been about to leave the washing-cabinet which divides the two compartments, and had taken the wrong direction and come into my compartment by mistake. Jumping up with the intention of putting him right, I at once realized to my dismay that the intruder was nothing but my own reflection in the looking-glass of the open door. I can still recollect that I thoroughly disliked his appearance.[1]

What Freud describes here is an incident of recognition, though one which is disruptively predicated upon the radical loss, or *defamiliarization*, of the known contours of personal identity. Alone in his railway compartment, the traveller fails to recognize his own reflection in the looking-glass. At once he realizes his mistake, but he is left with a sense of alienation caused by his distaste for the image he has seen of himself. A trivial enough experience: though nothing, we know, is trivial in Freud.

There is also something unsettling about the apparent inconsistency of the record. Freud says the intruder 'came in'; but he was himself sitting down, and only *after* the intrusion does he get up to evict the stranger. How could he have made the mistake he did if he saw only the reflection of a seated man? There is an instant in this account where the two images, of sitting and standing, seem to be transposed. Freud's own consciousness of the event, and his memory of it, produce a composite picture where the intruder 'comes in' and the traveller gets up to evict him in a single sequence which is referred to the narrator's own logic of action. Yet a simpler

phenomenal logic would suggest that Freud got up when the door swung open, saw his reflection in the glass, assumed that an intruder had come in, and so on.

Freud himself uses this incident to theorize the nature of the 'uncanny' (*das unheimliche*) in human experience, arguing that 'the uncanny is in reality nothing new or foreign, but something familiar and old-established in the mind that has been estranged only by the process of repression.'[2] The entire structure of Freudian psychoanalysis is built upon this premise of the new as the familiar but forgotten. Intimate self-knowledge is the possibility of the self's recovering what it has estranged through repression. This is an important cognitive model, but for the moment I will use it only as a point of entry into the problem of recognition that is my central concern in this paper.

Recognition is one of the oldest terms in aesthetic theory, employed by Aristotle to define a crucial moment in the epistemology of the literary text. The word Aristotle uses is *anagnôrisis*, which means a recovery of something known, rather than (as Aristotle himself explains it) a simple shift from *agnoia* (ignorance) to *gnÙsis* (knowledge).[3] The term's etymology requires it to be translated by the English word *recognition*, which stresses, in however muted a form, the element of repetition in its structure. It is true that Aristotle himself does not directly emphasize this: but he links the term with the *peplegmenos*, or 'folded' (usually translated as 'complex') plot, and with the associated process of *peripeteia*, or reversal. In the context of tragedy, which is where we find the three terms in conjunction, *anagnôrisis* is part of that movement by which the complex plot, folded upon itself like a piece of pleated paper or a skein of wool, can be opened out or unravelled through a reversal which is also a reversion or recourse, returning us to the beginning of the story and requiring us to re-read it in the light of what we now know. This knowledge comes to the protagonist, then, in the form of a paradox: it was always there in experience, yet it must be recovered from the hidden structure of events.

In classical tragedy this structure is best exemplified by *Oedipus Rex*, whose hero is compelled to recognize in his own person the contrary identities of being chosen and being accursed. He is ignorant neither of his destiny nor of the events in which he has participated; the discovery he makes is that these two are the same, and that his identity is such that he has always feared. What the

oracle has told him is not an *end* that he can escape, but a truth *constitutive* of his being. In a sense, therefore, the extension of plot in space and time, suggesting the possibility of travelling away from, or fleeing, as Oedipus tries to do, a predetermined structure, is countered by the collapsing of the narrative back into itself through Oedipus' enquiry into causes and origins. The retrospective movement of the plot in *Oedipus Rex*, a feature that led structuralist critics to see it as the paradigmatic detective story, makes recognition and reversal inescapable parts of the hermeneutic circle. Oedipus cannot know what he is if he did not, in some sense, *already* know what he is. This is of crucial importance to the Freudian plot of the return of the repressed: recognition is here characteristically reflexive, not directed at objects or actions outside the subject or knower, but at 'those wishes, repugnant to morality'[4] which are the subject's psychic property. The fictional structure of *anagnôrisis*, which theorizes a certain kind of relation between the self and the world, is collapsed into a form which posits only the problem of the self. In consequence it becomes easy to read tragic recognition primarily as self-recognition, ignoring the numerous and complex instances where the recognition scene involves the discovery of actions and motives belonging to others. The strictly reductive schema of the Freudian plot, similarly, makes knowledge *by* a self in the end no more than knowledge *of a* self.

Gilles Deleuze relates recognition to that movement he describes as repetition, through which the tragic hero *re-presents* a given knowledge to himself:

> This knowledge is in him, it is immersed in him and acts in him, but acts like something hidden, like a blocked representation . . . this unknown knowledge must be represented as bathing the whole scene, impregnating all the elements of the play and comprising in itself all the powers of mind and nature, but at the same time, the hero cannot represent it to *himself* – on the contrary, he must enact it, play it and repeat it until the acute moment that Aristotle called 'recognition'. At this point, repetition and representation confront one another and merge, without, however, confusing the two levels, the one reflecting itself and being sustained by the other, the knowledge that is represented on the stage and as repeated by the actor then being recognised as the same.[5]

Would it be right to say that the truths literature offers us are all of this kind – that is, they are truths we already know, and are forced to re-cognize by the process of re-enactment? I am reminded here of a story told of Paul Dirac, who said once to a colleague: 'How can you do both physics and poetry? In physics we try to explain in simple terms something that nobody knew before. In poetry it is the exact opposite.'

For Dirac physics offers *new* truths, poetry old ones: the first is a form of knowledge, the other of repetition or *re*-cognition. Dirac's distinction privileges the epistemology of the natural sciences, and this is a privilege we are normally inclined to concede. Our sense of knowledge itself as a corpus is derived from faith in its empirical content and in the capacity of the mind to acquire new knowledge from observation, deduction, and speculation. This is linked to the idea of progress which is fundamental to European science, positing knowledge as a process of growth and accumulation. In an essay published in 1911, the phenomenologist Edmund Husserl noted that the achievements of the natural sciences had been so impressive as to lead to a crisis of confidence in the humanities.[6] This crisis is one we are still facing, and it has led to an obsessive absorption in problems of textuality and interpretation.

Yet even for the sensory realm, it is difficult, as Peirce saw, to abstract what Descartes would call an intuition from the principles, even the semiotic tools, that are required to organize it in our understanding. Peirce stressed the inferential character of perceptual judgments, and inference requires not only a hypothesis, but a form of corroboration (which may of course be gradual and incomplete) based on previous knowledge.[7] Kant considered space and time as forms of intuition (*Anschauung*) which make our perceptions of a certain nature, but have no necessary relation with the nature of things themselves. Things themselves are unknowable, and Kant was interested not so much in the empirical problem of how we come by our *knowledge of* objects, as how we establish our *knowledge that* certain propositions about the world are true. In consequence, as we know, he was concerned to elaborate not on the intractable matter of the world, but on the *a priori* concepts, or categories, by which our knowledge is ordered and realized. Beyond the categories lie the transcendental Kantian schemata, which, as Peirce says, if Kant had considered them early enough, would have overrun his whole work.[8] Strawson makes an analogous point when commenting on Kant's 'apparently technical' use of the term *imagination*, which

implies that:

> Recognition of an enduring object of a certain kind *as* an object of that kind, or as a certain particular object of that kind, involves a certain sort of connexion with other non-actual perceptions. It involves other past (and hence non-actual) perceptions, or the thought of other possible (and hence non-actual) perceptions, of the same object being somehow alive in the present perception.[9]

(There is a similar Wittgensteinian argument regarding 'seeing' as 'seeing as'). And yet Kant does not by any means deny the existence of the world, whose phenomena he accepts by a kind of 'naive realism' which remains in the main undefended.

First knowledge of that world, that is, how *something new* is known in its newness, might be described as the hard problem of cognition. If the finite intuition, as Heidegger said, is 'in the first place anchored to the singular',[10] how do we see something (say a colour) for the first time before we compare it with something else? This problem is related to the question of *what it is like* to see a colour, and the difficulty, which seems to me crucial, of determining whether this is an epistemic question or a description/replication problem.[11] It is a problem of the phenomenal consciousness, which possesses intuitions – the redness of red, for example – neither replicable nor communicable in their entirety. These constitute a body of experiential knowledge which is not re-knowable by others, nor analysable by the mind *except* through forms of abstraction and correlation. We could argue that the empirical intuition exists, but never constitutes knowledge as such. For that end it must be integrated into a cognitive scheme, operating through the principles of identity and difference. Are these principles axiomatic? Moreover, what would be the difference between knowing that something is a tree (even a new, previously unobserved kind of tree) and that somebody is the same person we saw yesterday? A further problem arises if we ask, *who knows?* and *who knows what?* When Descartes says of the piece of wax: 'It is of course the same wax which I see, which I touch, which I picture in my imagination, in short the same wax which I thought it to be from the start',[12] two premises are involved – first, that of the unity of the thinking subject, and second, that of the collaboration of the faculties in establishing the identity of the object. Recognition is insufficient in itself to serve as a model of thought, though it is built into the structures of perception.

Another model is presented by the Platonic notion of *anamnêsis*, which makes all knowledge a recollection of what the immortal part of our being already knows through its identity with the absolute. This belief was central in the enterprise of the early modern scientific revolution (in the thought of Galileo, for example) and is echoed by some modern scientists at least. Philosophical Platonism is very often a part of the cognitive scheme assumed, for instance, by mathematical physics.[13] This is not in the least Kantian, since the truths Plato speaks of are not the product of our own modes of thinking, but exist independent of us. They constitute an absolute truth-content to which our ideas correspond and which not only confirms our speculation, but which we *re-cognize* when we glimpse a mathematical truth. Even those scientists who would not subscribe to this belief might have difficulties finding a way out of the hermeneutic circle which the notion of truth itself implies.

Yet the hard sciences have made extraordinary gains, as Husserl saw, by bracketing off these problems of cognition (a disappointment to a philosopher who was particularly interested in how the mind acquires its *knowledge of* phenomena) and assuming that knowledge of singulars takes place, that it is cumulative and that it can be organized. Certain kinds of philosophy too have benefited by assuming what Quine called the 'inscrutability of reference': that is, by bracketing off the world and its objects from the domain of philosophical analysis. It is not my intention to worry indefinitely here about these problems of knowledge and truth. Some 'unexamined' assumptions work perfectly well for the real world. Physics is therefore a branch of knowledge, though we have not worked out how much re-cognition is involved in its practice.

Nevertheless, if we now look at the other side of Dirac's distinction, we may feel that poetry, apparently a poor relation of the hard sciences in epistemological content, seems to offer some truths after all. The recognition plot, using *anagnôrisis* in the classical, or Aristotelian sense, is only one example of a confirmed identity. Tragedy employs such structures to undermine our hope in the possibility of getting away from ourselves and our actions. As Aristotle stressed, *anagnôrisis* confirms causal necessity: it is linked to the probable (*to eikos*) and the necessary (*to anangkaion*). Life overwhelms us in difference, in the chaos of particulars which occupy space and time. When, in the midst of these, we recognize others or ourselves, we make what we might call a *human* relation, by an intellective act

which is the basis of knowledge itself.

This act involves a certain abstraction and generalization which is built into the hermeneutic circle. In rhetorical terms we would describe this act as the figure of synecdoche. The plot of recognition, whether in tragedy or in the contemporary detective story, is normally dependent upon signs and tokens of presence. Classical writers call these *gnôrismata* or *sêmeia*, objects indifferent in themselves, which can nevertheless function as means of knowledge to the person who recognizes their significance. The process involves inference as well, either in the true deductive mode of syllogism, or in the speciously analogous form of paralogism (*Poetics*, 16). Specific correspondences tell Electra that the stranger she sees is her brother Orestes; a number of clues prove to Holmes that Silver Blaze is himself the murderer of his corrupt trainer. The narrative urges us that the part implies the whole; at the same time, it shows us that if a general premise is valid, its particulars must be assumed to be true. This is a syllogistic, or – as Aristotle, again, described it – a *paralogistic* structure, having the appearance of logic without necessarily satisfying the strictest logical criteria.

Anagnôrisis is not peculiar to tragedy: it is a feature of many narrative plots, a means through which the narrative 'makes sense of' the world by grouping exemplars in what we might call a similarity set. Similarity is understood as the obverse and cognate of difference, both necessary components of our experience. Not all experience, of course, will provide the radical and unsettling doubling of identity which disturbed Freud in his railway journey. Like the myths of psychic identity that Freud himself uses – such as those of Oedipus or Orestes – the story of the mirror in the washing-cabinet is an extreme example of its kind. Freud was fond of such examples, by which he resolved alterity into sameness by recourse to what we can only call *poetic truth*. Something of this effect is to be seen too in his famous analysis of Wilhelm Jensen's novel *Gradiva*, with its extraordinary superimposition of archaeological trace, work of art, memory, dream and reality.

That recognition is built into the structures of literature and art does not mean that art is a mirror. Just as knowledge (*pace* Rorty's argument)[14] is not a mirror of nature, but constructs its own synthesis by means of the intellect, language makes its meanings accessible to us through a process of endless semiosis where actual objects are never touched as such but always converted into signifying forms.

Linguistic meanings, as we all know, are produced in a web of sameness and difference. The differential structure is only the ground from which the recognitions that constitute a signifying term are generated. Language imposes a process of recognition upon us along the signifying chain, so that grammatical syntax can be seen (as it has often been seen) as analogous to narrative.

Not for nothing, then, is *anagnôstes* the Greek word for a reader, or for a slave trained to read. Plato uses the term approvingly for Aristotle himself, as Renaissance commentators are fond of noting.[15] Reading is pre-eminently the act by which we make the connexions that our interpretation of the world requires of us. Literature asks us to read, but so does the world. In literature, however, perceptions reach us in an endlessly reduplicating medium, that of representation or (to use the classical term) *mimêsis*. In Plato's famous reductive argument (*Republic X*), mimesis is trapped in the same kind of secondariness that Dirac attributes to poetry. This is a variant of the mirror analogy, which Plato also uses: according to this argument, literature reflects reality, and we may indeed see our faces in it, but we learn nothing from the experience, and the image itself may be distorted. Literature is therefore not a true form of knowledge, but is likely to mislead us by its simulations of non-existent identities (like Freud's reflection in the looking-glass).

Yet in fact the stranger Freud saw in the glass was recognized, not without some disaffection. The imagined other, produced by a logical and perceptual confusion, was resolved into an aspect of the self. As an exemplar, the anecdote serves also the neat theoretical purpose of illustrating the thesis of repression. It is thus a double in another, semiotic sense. Indeed it seems to combine the three functions of Peirceian semiosis: the story is an icon which persuades through resemblance, it is an index which serves as a clue, and it uses the symbolic medium of language to make a point about phenomena in the world. Its mimetic truth offers itself for re-theorization on Aristotelian lines as an abstraction of universals from particulars.

It could be argued, thus, that literary representation, with its pervasive use of the model of recognition, confirms the ubiquity of that model in thought, which represents the world to our understanding. One version of representation ceaselessly rids the world of difference, or rather, domiciles it in the house of the Same: '*difference becomes an object of representation always in relation to a conceived identity, a judged analogy, an imagined opposition or a perceived similitude.*'[16] This is not the version I wish to adopt. It has been my argument in this paper that we cannot distinguish in mental life between the primary and the secondary, between cognition and recognition. Thought is inescapably reflexive, returning upon itself in the very moment of recording a perception. The recognition plots of literature, like language itself, foreground the endless re-presentation of experience in signs and in forms of knowledge. The literary text is necessarily founded on this structure of loss at one level for recovery on another. In one respect it offers a fiction of reassurance; in another it seems to defer cognitive pleasure, collapsing the unknown into the always-already-known. The process of recovery and re-enactment which literature accomplishes is also mimetic; it simulates the extent to which re-knowing, in the Western philosophical tradition, is implicated in the act of knowing. Like knowledge systems, literary texts are constructions of memory, binding together the insistent particulars of experience by what Henry James called 'the spell of recognitions'.[17]

Should we conclude then that the literary model of recognition is only a means of confirming known identities? This is not my conclusion. If poetry, as Dirac thought, tells us no more than what we always knew, it also tells us what no one knew – or recognized – before. It was only because Freud had never seen the stranger he saw in the glass that he 'thoroughly disliked his appearance'. It is this instinctive moment of revulsion that constitutes his knowledge of the episode: the *othering* of a self presumed as familiar and recognizable. In an analogous way, literary texts thrust difference upon us, not simply in respect to ourselves, but in relation to others. They are constructions of other minds which compel us to accommodate them to our own. The experience of reading is therefore as much that of encounter as it is one of repetition; indeed, if repetition were the only meaning of recognition, we would be condemned endlessly to repeat the world in thought. By contrast, thought shares in this and other worlds, just as literature too offers us possible worlds different from our own. Iser explains this in his theorizing of the fictional in its 'boundary-crossing' capacity: 'mimesis is a repetition which produces something'.[18] When we turn back to Freud's anecdote of his experience in the railway-carriage, we may be struck afresh by its *staged* quality, the extent to which it produces, like a literary text, the shock of the new as well as the recognition of the same. In terms of the text which incorporates them, these are

purposive fictions. They remind us of the strangeness and intractability of the world, returning us to a sense of the self as an other among others.

The fact that literature does this is obviously related to the fact that experience, too, convinces us of the insufficiency of the recognition model in and for itself. This is well argued by Deleuze:

> Something in the world forces us to think. This something is an object not of recognition but of a fundamental *encounter*. What is encountered may be Socrates, a temple or a daemon. It may be grasped in a range of affective tones: wonder, love, hatred, suffering. In whichever tone, its primary characteristic is that it can only be sensed. In this sense it is opposed to recognition. In recognition, the sensible is not at all that which can only be sensed, but that which bears directly upon the senses in an object which can be recalled, imagined or conceived . . . The object of encounter, on the other hand, really gives rise to sensibility with regard to a given sense. It is not an *aisthēton* but an *aisthēteon*.

Deleuze is deeply hostile to the recognition model, which he connects to the Platonic model of reminiscence in offering a transcendent theory of knowledge. This response too seems to me to be founded on a misunderstanding of what recognition and, by extension, representation, involves. We think of representation as being *by* and *for* consciousness, *of* the phenomenal world. But in fact we live in a world constituted by recognition and representation, not one which can be bracketed off from the consciousness which experiences it. The texts of the world, like those of art, present us with strangers who may also be our sisters – or our mothers, as Oedipus found. The force of recognition is brought home in the compulsion to acknowledge this relation, even where it seems most alien and most unwanted. What we gain thereby is not a simple confirmation of what we already knew, but a more difficult and more radical knowledge. When we ask, *what is recognized?* we may answer that it is not the same but the other, the world that is not ours alone to experience or to think.

Literature founds this knowledge on an experience of the fictional: which is not, for that reason, simply the untrue. Poetic representation stages the world to our consciousness, as well as our consciousness to the world. The act of knowing in this double staging can only be seen as an act of recognition, which is why the scene of *anagnôrisis* is proximate to a reversal, or doubling back of the plot on itself. But it would be wrong to see this moment purely as the doubling back of consciousness (in the narrow sense). Rather, the moment is one in which consciousness receives the existence of others into itself, accepting the burden of difference as well as that of relation. In this respect the literary text impels us not to solipsism but to charity. That we can read at all would argue that language and the world are shared, though literature must be above all the instance that convinces us that there are other ways of thinking the world. At the same time, it must be also the instance that proves that these other ways are open to our imagination and understanding, thus confirming the Davidsonian thesis of shared meanings rather than the poststructuralist hermeneutics of suspicion.[20] The literary text takes us out of ourselves at the very moments when we believe ourselves to be most alone; it assures us that we are not only as we see ourselves, but as others see us. In classical tragedy the moment of this recognition is also a moment of strangeness or *wonder* (*to thaumaston*) and it is this that constitutes the pleasure (and implicitly the profit) of poetry.

I would like to conclude by reflecting on two signal moments of recognition in literature, each canonical in two quite different and separate literary traditions. The first is the recognition of Sakuntala by Dusyanta in Kalidasa's play *Abhijñāna Śākuntala*. It is a moment hedged around with miraculous and necessary provisions, especially the pre-recognition of the ring which has restored the king's memory. The encounter itself is elaborately prepared for, not only expected but anxiously desired by the audience, telling us no more than 'what we already knew'. But the lovers are taken by surprise, and the grief of the radical forgetting that had separated them is present in the king's amazed recognition of Sakuntala's continued identity through altered appearance.[21] Wonder functions here as a reminder of what we may lose through forgetting the existence of others in the world. The king's memory does not simply confirm 'what he already knew', it makes his own identity continuous with a larger objective reality he had in the past denied. In my other example, the recognition of Cordelia by Lear in the fourth act of Shakespeare's play, it is uncertain whether this larger objective reality is present to Lear's afflicted consciousness at all. In this condition, his ability to recognize his daughter is itself a wonder in a world beset by tragic accident:

I am a very foolish, fond old man,
Fourscore and upward, not an hour more nor less;
And to deal plainly,
I fear I am not in my perfect mind.
Methinks I should know you and know this man,
Yet I am doubtful; for I am mainly ignorant
What place this is and all the skill I have
Remembers not these garments; nor I know not
Where I did lodge last night. Do not laugh at me,
For, as I am a man, I think this lady
To be my child Cordelia.

(IV.7.60-69)[22]

Lear's act of recognition is not primarily an act of remembrance at all. Cordelia enters his consciousness as someone who is not contained by it as memory, who exceeds the structures of recollection and identification already, as he says, so much impaired. The force of this meeting is the force of encounter: only afterwards can Lear think of the past with guilt:

I know you do not love me, for your sisters
Have, as I do remember, done me wrong.
You have some cause, they have not.

(IV.7.73-75)

But at the moment of seeing and knowing Cordelia, he knows her as someone other than the things he may or may not recollect: someone known, and therefore recognized, but also someone encountered, and therefore known for the first time. It is this profound coincidence of experiencing the new and recognizing the same that literature struggles with in its effort to tell us *what an experience is like*. This may indeed be why Freud chooses the anecdotal, literary mode of telling in his recounting of an experience of recognition.

NOTES

1. Sigmund Freud (1919), 'The Uncanny' in *The Standard Edition of the Complete Psychological Works of Sigmund Freud (SE)*, trans.introd. James Strachey and Anna Freud, asst. A. Strachey and A. Tyson, 24 vols (London: Hogarth Press, 1953-74) vol. xvii, p. 225.
2. *Ibid.* p. 221.

3. Aristotle, *Poetics* 11.4.
4. Freud, *The Interpretation of Dreams*, *SE* iv.263.
5. Gilles Deleuze, *Difference and Repetition*, trans. P. Patton (London: Athlone Press, 1994), p. 15.
6. 'Philosophy as a Rigorous Science', in E. Husserl, *Phenomenology and the Crisis of Philosophy*, trans. Q. Lauer (New York: Harper and Row, 1965).
7. 'Some Consequences of Four Incapacities' (1868), in *The Writings of Charles Sanders Peirce*, vol. 2 (Bloomington: Indiana University Press, 1983), p. 213. See also Peirce, *Collected Papers* (Cam. Mass: Harvard University Press, 1934-48), 5. 116.
8. Peirce, *Writings*, 5. 258-59. I am indebted here to Umberto Eco's discussion of these problems in *Kant and the Platypus: Essays on Language and Cognition*, trans. A. McEwen (London: Vintage, 2000), esp. chapters 2 and 3. See also Paul Feyerabend, *Against Method* (London: Verso, 1978) Introduction and *passim*.
9. P. F. Strawson, 'Imagination and Perception', in *Experience and Theory*, ed. L. Foster and J. W. Swanson (Amherst, Mass.: 1970), p. 43.
10. Martin Heidegger, *Kant and the Problem of Metaphysics* (Bloomington: Indiana University Press, 1973), p. 2.
11. See Thomas Nagel, 'What is it like to be a bat?', *Philosophical Review*, 83 (1974), pp 435-50, and David Chalmers, *The Conscious Mind* (Oxford: Oxford University Press, 1996).
12. René Descartes, 'Second Meditation' in *The Philosophical Writings of Descartes*, trans. J. Cottingham, R. Stoothoff and D. Murdoch (Cambridge: Cambridge University Press, 1984), vol. II, p. 21.
13. See, for example, Roger Penrose, in *The Emperor's New Mind* (London: Vintage, 1990), pp 147-50, 552-54, and *Shadows of the Mind* (London: Vintage, 1995), pp. 412-17.
14. Richard Rorty, *Philosophy and the Mirror of Nature* (Princeton: Princeton University Press, 1979), 3.3.
15. See, for example, Pico della Mirandola, *Oration. On the Dignity of Man*, in *Renaissance Philosophy*, I. *The Italian Philosophers* ed. A. B. Fallico and H. Shapiro (New York: Random House, 1967), p. 160.
16. Deleuze, *Difference and Repetition*, p. 138.
17. Henry James, Preface to *The Portrait of a Lady*, introd. Graham Greene ed. Nicola Bradbury (Oxford: Oxford University Press, 1981), p. xl.
18. Wolfgang Iser, 'Towards a Literary Anthropology', in Ralph Cohen, ed. *The Future of Literary Theory* (New York: Routledge, 1989), p. 227.
19. Deleuze, *Difference and Repetition*, pp.139-140.
20. Donald Davidson, 'On the Very Idea of a Conceptual Scheme', 'The Method of Truth in Metaphysics' and 'Communication and Convention' in *Inquiries into Truth and Interpretation* (Oxford: Clarendon Press, 1984).
21. *The Abhijñānaśakuntala of Kalidasa*, ed. S. K. Belvalkar (New Delhi: Sahitya Akademi, 1965), VII.22, p. 136.
22. William Shakespeare, *King Lear*, IV.7.60-69, ed. R. A. Foakes (Arden, 3rd Series, 1997)

I am generally indebted to Terence Cave's stimulating study, *Recognitions: A study in Poetics* (Oxford: Clarendon Press, 1988).

Jaspers' Communicative Truth as a Creative Challenge

PAULA ARIZPE

'Communication is the way truth is revealed in time'.[1] For Jaspers, truth is communication understood not as the mere transmission of contents but as the fulfillment of one's own existence and the existential fulfillment of others. Thus Jaspers understands communication as a creative event by which reality is known through our own fulfillment.

Access to personal truth is conquered by means of the diverse perspectives under which reality is apprehended. Jaspers' point of departure of this existential and cognitive adventure is Kant's subject-object dichotomy. This fact splits knowledge into diverse perspectives therefore, the meaning of truth, at a first stage, will depend on the perspective under which the subject grasps the objects of the world. In the end, the different meanings of truth flow in unity above time, this is what Jaspers calls transcendence. The rational articulation of these meanings within a determined time and space fulfills *Existenz* – fulfilled subject in communication with the world – this enables him to develop an existential committment with truth as totality. For Jaspers, the fullillment of the self is the same thing as a radical committment to truth. This gives a trend to existence, guiding it toward transcendence. This is the challenge Jaspers' invites to: 'the belief in man's own possibility'.[2]

A methodology to openness

The author establishes that human wisdom goes far beyond scientific knowledge. His gnoseologic requirements go through the scientific world but they reach further because he expects to attain the true human realm in which authentic communication is feasible, that realm in which man can fulfill his own potentiality, at the same time he completes the human world unfolding his creative activity. It is a realm beyond objectivity, that Jaspers calls transobjective. It is the

situs where man can fulfill his own potentiality and it is in his own realization or the materializing of his personal possibilities that individuals are able to build an organic community.

We will therefore analyze the gradual steps into which the world opens to a genuine human profoundness considered by the author as the first condition for the development of human freedom.

The subject as *empirical conciousness* finds himself in front of a world of objects with which he may satisfy his basic needs. This relationship rests at a sensitive level, so at this stage truth holds a meaning of mere *utility* therefore is limited to the satisfactory-unsatisfactory category. Evidently it is a very poor perspective for man as a whole, therefore man experiences a deep dissatisfaction if he rests anchored to this infimal truth. He must therefore, continue forward in order to deepen his relationship with the world.

The subject as *general conciousness* is able to structure formal concepts of the world. By organizing them in such a way he may obtain practical results. This formalization opens a more immaterial perspective of reality so truth takes an *apodictic meaning*, moving through categories of consistency/inconsistency, in/out of the system. Up to this point communication is sufficiently precise so as to join members of a community around the achievement of material goals. Individuals can be exchanged indistinctively; therefore communication at this level is not humanly satisfactory. Following this approach to existential communication, the individual can go further in the communicative relationship.[3]

The subject as *spirit* is able to choose ideals. At this stage he faces decisions and discovers the realm of freedom, therefore he can drive towards the goals he selects after he defines his own orientation in the world (*Weltorienturiung*). This realm comes closer to the human realm so therefore truth takes the meaning of *clarity* and moves within the intrinsic/extrinsic category. Nevertheless, man as a whole still has to decide towards transcendence.

With these means, Jaspers understands man as possible existence to the depth of existential communication where truth takes the meaning of *belief in openness*. This openness consists in a sort of knowledge acquired by presence in profoundness which includes the distance of reflection and the immediacy of intuition as cognitive processes enlarded together. Openness is the only category that can harbour the human aspiration of creative knowledge which is as wide as freedom. Human truth is the belief in what we are not yet

but can become, a belief in what we have not yet done but shall do. Freedom is an adventure to accomplish that can only be conquered in the daily struggle to exist.

I am in search of transcendence once this profoundness has opened before me, and inside my temporal existence, searching becomes finding: because man's historical existence, understood as historical, can only root in unity with what is being searched for, that is to say, that presence that exists only as search, that which cannot be divided.[4]

When the objectual being is transcended, as it happened in the world orientation stage (*Weltorienturiung*) the object is not grasped as I grasp my own self. In the *Existenzerhellung* stage, I have a notion of the object due to internal action, that nevertheless its failure at grasping wholeness, rests in the authentic being. . . . The modes of this search of being at a possible existence level are always the paths of trascendence. Its explanation is always metaphysical, because this process moves in the transobjective realm of freedom.[5]

Thought in suspension or way to creativity

Jaspers' methodology is not a linear way that goes from one point to another. On the contrary, it pretends to emphasize circularly and tautologically the different aspects under which each phenomena of reality should be integrated and under which each perspective should be appreciated in all its magnificence. We should be "open" so as to let reality impress us in all its grandeur. Only by means of "thought in suspension"[6] can each object of knowledge become an encounter, that may make inchoate a creative event in our inner self. Every creative event links different aspects of reality which the scientific, or the modern philosophic approach tends to dissociate as "contrasting" or "dilemmatic", or "opposed". When a thinker acquires the habit of distinguishing the diverse aspects of reality and of admiring them in the multiplicity of its of its possible relations without running into insurmountable scissures, he gains a surprising freedom to understand the creative phenomenon from its genesis. The creative phenomenon is that in which different aspects of reality and temporality coexist by the power of the *Existenz*. To harmonize all the relevant aspects of reality it is necessary to tense the mind and adapt it to thinking in suspension.

The vinculation of this powerful presence as transcendence in

which all reality that is relevant, is sustained – imposes itself discretely as a supreme and absolute instance. This instance is understood as infinity or the unconditioned or free existence, it is a search thought and cherished by the most intimate action of the subject which gives us the idea of the circular movement appeal-response, or the nutritious scope of reality which characterizes Jaspersian metaphysics. This profound reality, genuinely human, needs to be understood within the active-receptive categories which are the means in which the dialectic of game develops. This is the creative reality, appealling and defying, more palpable and vital to which the author invites.

His deep knowledge of the logic of creativity enables us at the same time to apprehend the special details of the logic of participation and of encounter, taking us directly to the discovery of transobjective realities.[7] Objective realities are those by which the existent orients himself in the world. These are objects measurable, delimited, verifiable, and quantifiable by everyone. Transobjective realities on the other hand, are those which have a superior spacious-temporal status. More than than that, they and cannot be restricted or located. They can only be grasped by an existence in its full exercise of freedom. As they cannot be located at a purely cognoscent's distance – because they commit him who wants to grasp them – they englobe him, because they offer the possibility of deep communication: that of game. They are not objective as they cannot be projected. They can only be apprehended by the type of participative knowledge that links knowledge and love, knowledge and creative action, knowledge and active-receptive immersion in relevant realities that are open to the foundation of ludic spaces. The transobjective world where nothing is "given", where everything has to be done in a personal and creative manner, can only be accessed by "leaping into the void".[8]

Participative ludic scopes

Creativity is a relational phenomenon enlightened by the free scope of game that is established between the *Existenz* and the comprising realities with which it builds a committed encounter relationship. Understood from this creative point of view, the event of game means the active-receptive immersion in realities that offer a multiplicity of aspects that can be integrated such as shape and depth, grace and dignity, movement and stillness.This differs from the Kantian

proposal that tends to dissociate the diverse elements and moments of experience. Jaspers seeks a profound integration transcending the object-subject scissure through creative praxis. This integrative practice is what *game* consists of: a fertile interrelationship of events endowed with meaning, which in coming together, open the sense to the plentiful existent. In this way, truth is an enlightment that happens within a relationship of encounter, where the object not only manifests its form but unfolds its interrelational possibilities. This conception of reality projects the basic parallelism of the human being: matter as well as spirit. True humanity is a happening that takes place in the transobjective scope of creativity where both elements — matter and spirit — can be joined harmoniously, freely and renewedly. Creativity is the impulse to take part in the game. Game is that which is not contingent, neither in the objective nor subjective sense, nevertheless it is not coercive neither externally nor internally. That which is human or existential requires for its fulfillment and perfection to open itself to exterior reality with all its norms and rules (something not contingent but necessary). The acceptance of normativity is not coercive when *Existenz*, endowed with the will of co-creation finds a rainbow of possibilites of action. He then chooses and builds, that is to say, he makes them his. This type of creativity turns graceful, luminous and expressive when both flows of reality, subject and object, interrelate in a ludic manner forming a unique scope: an expressive game.

Integration means freedom

As we have seen, *Existenz* conforms as spirit departing from objective reality, joining objects together from a higher perspective, more freely and serenely. When empiric truth connects with formal truth, it gives way to the law of necessity; not only it is not tensed by dissociating abstraction but it is accompanied by the immediate intuition of creativity. This dialectic in a comprehensive horizon or scope of immersed realities is parallel to the dimension of intuition and discourse, if by intuition we understand the human capacity to penetrate in an immediate way such realities endowed with meaning. And if by discourse we understand paying successive attention to the different perspectives these realities offer, a mutual potentiation can be observed between intuition and discourse. This mutual

potentiation of discursive mediation and immediate intuition take us to the realm of intellectual intuition being immediate and also indirect and reflexive. This original conception is punctualized by the mutual implication of historicity and truth, both central themes in Jaspers' thought. All along, these medial elements are assumed by the Existent who has developed the creative impulse and are turned into transparent vehicles of the presence of the reality which they mediate for. This proposal emphasizes the dynamic-relational character of truth and of the cognitive process in general, making the encounter with reality as a creative challenge possible.

Nothing is more intimate to man than that which enables his plenitude, that which contributes to his own existential fulfillment and the impulse towards it. In this sense, objects are penetrated by the *Existenz* in a directly proportional way as he lets himself be penetrated by the objects of reality in all their depth. This is a mutual potentiation based in the answer-response scheme or circularity of the game – game, that is, to be immersed in an englobant reality which implies answering positively to a high and appealing instance. In other terms, when in a community of ideas, the realm of reality becomes familiar or inhabitable, it loses seriousness because it takes on its convenient size. When it is related to sensation, necessity loses its opacity as it turns clear and simple.

Because of its integrativeness, the game adjusts and adapts itself constantly and therefore potentiates the mutual opening of creative possibilities. To create, weighing the different possibilities that reality opens to Existence is authentic freedom. Seriousness closes possibilities. To play implies "perform within the game" our possibilities of action: and this is to exercise freedom.

The ludic activity surpasses the object-subject scissure and this is an important fact because it incites a serious commitment with reality. The authentic seriousness of a game comes from the 'inside', from the game-players'dialectic, from the active-receptive immersion of the player in the scope of possibilities that each game opens with its peculiar normativity. The game is serious because it encompasses, subject and object are unity, in an actual time and place. This constitutes truth as totality grasped by *Existenz* in time: the game fulfills the existent while it is played. It is clearly not an object, because what game is, cannot be grasped by a speculative intention. Knowledge of this sort can only be acquired in the profoundness of

the ludic experience. The player is a scope full of possibilities of action ruled by ludic norms, and the game is a scope of possibilities of action with a purpose that transforms the player in all his entirety. The ambital-relational character of the ludic experience manifests its fecundity to achieve creative praxis. Reality is not closed in itself, it is precisely an open field for interactional creativity.[9]

This is the meaning Jaspers tried to describe with the term "possible existence": a human being who is never sufficiently fulfilled, who tends to open his scope of reality and who is called to unfold his range of action in all directions. Therefore, conceptualized as a totality, the human being is an ambital being, open to develop himself by the constant foundation of communicative relationships or scopes of game.

To inhabit: unity in diversity

In this way, existence conforms itself with the creation of scopes. This creative enterprise supposes the permanence that "inhabitance" implies. Everything valuable that man fulfills is due to permanence, and for staying it is necessary to root in good soil, that is: to inhabit. The human creation of scopes – scopes of knowledge, friendship, love, art, etc. – runs parallel to the creation of physical places that in some way are expressions of the human dynamism which inhabitance consists in. These spaces are genuinly inhabitable when they can be recreated by persons who live them daily. How is this possible? There are spaces in which man finds himself "at home" where he feels a subtle adequation – absolutely perceptible and hardly analyzable – that is given between his spiritual conditions and the sorroundings. Certain old cities, plazas, domestic atmospheres, sacred places provoke in man a pleasant feeling of balance and harmony. This harmonic coordination is produced when man recreates in his quotidianity the spaces that structure the train of his daily existence and in this way he turns a physical space into a true human scope. Real space, the only one genuinely human, is that space carved by a community through its daily living. There is nothing strange in feeling a comfortable impression of peace – which is the fruit of fulfillment – when visiting some old cities created slowly by their inhabitants, developed organically from the inside to the outside, just conforming to their intrinsic requirements or their own logic. Space represents in a community's life an analogous role to time in

a symphony: both build a natural and necessary unfolding scope, a horizon in which man can expand his possibilities.

These scopes are not "given" to *Existenz* naturally; The existent has to found them freely within the tense unfolding that life in community implies, because vital space is not static – as it is commonly thought – it conforms itself by a very complex weave of dynamic interrelations and internal tensions. Possible existence should decide and thus fulfill itself to become *Existenz*: this is the authentic realm of freedom. The *Selbst* (actual being-himself) relates with *Selbst* (future being-himself) in a creative way: this is the human challenge. This frame of interrelations constitutes an order, an ordering, a purpose. The order that ought to be learned and revived by man that inhabits a space, if this space has to be rigorously inhabitated, has to produce balance as a feeling of peace without provoking a dissolute relaxation. It should enable a profound integration of the spirit. Consequently, any study about space transforms itself into a praise of order, positively understood as an eminently active ordination. It is at this point that the virtuous cycle conforms itself: the dynamism of human existence confers an ordination and a sense of space. The virtually structured space makes a sense or meaning of life possible. Such is existence when it moves inside a determined space in the intertwining of human relationships. For the author, this is the reason why man inevitably loses himself in the shapeless magma of a limitless world (empiric existence), lacking the sense that only creative immersion or existential delimitation confides. Existence cannot live statically or passively in a space. The existent has a primal need to install himself actively in an encompassing scope, configurating, and communicating himself with all the dramatism its existence implies. In this way, the author tries to delimit the scopal reality of existence.

Jaspers pretends to delimit the ambital reality of existence because this is the only adequate *situs* where it can achieve its own posibilities: making use of its freedom in the encountering with existence. This encounter with oneself is a creative contact at an essential level. Man is primarily inclined to the foundation of new scopes in collaboration with other beings. These scopes are as strong and meaningful and as profound as are the beings that relationally conform them. The closer these are to their origin, the more "themselves" they are. This is the way the scope of freedom is enlarged by existence. Therefore, the creation of dialogic scopes is

the primary and highest activity of existence: the human being becoming such. Man is positively spatial because he has the ability to create interrelational scopes, for which physical unfolding represents only the material and obvious platform from which to build or conquer existence. The scopes are the result of the configurative action of the world, understanding the world as an open field of possibilities not as a mere physical space.

Jaspers clearly stated that empirical existence is only the substructure of possible existence. And possible existence becomes *Existenz* in the daily struggle to lead, ordaining its freedom towards transcendence, that is to say creating scopes where he fulfills himself. Therefore, the dynamicity of human existence is oriented towards the creation of scopes. For the author, to exist is to exercise freedom, and freedom in an absolute sense is communication.

Truth as a creative praxis

Jaspers describes in the *Philosophic Faith* and *Philosophie* how the leap (*Ursprung*) to transcendence is the centre of the fulfillment of possible existence. This leap that means the supreme and radical immediateness with reality is the more eminent form of existence because it is not 'given' as the *Dasein* is. The author shows that the highest forms of encompassing reality are those that require communication from the Existenz through the most vigorous forms of creativity, in order to found the most perfect genders of immediateness. These forms of immediacy are only possible in the core of a presential and encompassing closeness through a confident search. If man, in the search for security, holds on to objective methods adequate only for objective realities he makes transcendence dissappear.

Thus, truth is an unavoidable event. Truth is immersive, cocreative, ludic, it is an impulse to active fulfillment. Truth is encompassing.

The comprehension of truth is not a mere interpretation tecnique, but a dialogical and interactional praxis in which the existence commits freely with reality and its scopes. Man is open beyond what is given. That is why, there is a circular and creative relationship among our comprehensive activities and our openess towards totality: we only have to dare ourselves to exist creatively.

REFERENCES

1. Jaspers, K. *La Fe Filosofica.* Buenos Aires: Losada, 1968. p. 43.
2. *Ibid.* p. 5.
3. 'Philosophy tries to grasp the sense of life beyond all the means of this world, to manifest the significance these means to englobe; fulfill realizing in the present this sense, as if we said transversely in life – with its own presence to serve its own time and the future – never to diminish man or any person to the condition of a mere means.' *La Fe Filosofica,* p. 136.
4. 'At all times the problem has lied in the tension: to find independence out of the world, in renounce and solitude – or inside the world – through the world cooperating with or succumbing to it. Thus, the philosopher only wants his freedom along with the freedom of his fellow men, a life, only in communication with other men.' *La Fe Filosofica,* p. 137.
5. *Philosophie.* Vol. III. Berlin: Springer, 1932. p. 3. Free translation.
6. *Denken in der Schwebe.*
7. "Since the beginning there is something unmovable in philosophy. Every change in the stages and missions of the existence of humanity, in the progress of science, and the development of categories and methods of thought what is pretended is to grasp the one eternal truth with new methods, with possibilities of clarity perhaps broader. Our mission is actually to measure this eternal truth in nihilism taken to its futhest extreme. This requires from us the appropiation of tradition in a way that is more than contemplative, but that situates in its core as if it were its own cause.... What matters is the appropriation of the acquired perfection attained in any time by philosophy in order to keep communication incesantly renewed with the great phenomena of the past, not as overcame but as present." *La Fe Filosofica,* p. 142-144. Free translation.
8. *Ursprung.*
9. 'The disposition for an unlimited communication is not the consequence of a kind knowledge, but a resolution to follow the path of being human. The idea of communication is not a utopy but faith'. *La Fe Filosofica,* p. 146.

What is Truth? Art and the Mind

RIMI B. CHATTERJEE

Those who quote the most famous chiasma in the history of English literature often fail to notice or consider the quotation marks in which it is placed. "'Beauty is truth, truth beauty'" in John Keat's *Ode on a Grecian Urn* is followed by a dash, after which the poet completes his apostrophe of the Urn by saying, "That is all / Ye know on earth, and all ye need to know." For a piece of pottery, such a simplistic view of the world is probably appropriate, but naÔve readers of Keats do him an injustice when they equate this creed with Keats's own, or read this couplet as Keats's advice to the world.

The debate on art in Western society is coeval with the debate on truth. To find beauty in art is a matter of interpretation; Natural beauty exists independently of man, although we may disagree on our evaluation of different bits of it. But artistic beauty presupposes a creative process carried out by a human being; it presupposes an artist who is a fallible creature endowed with mind and will – in other words, someone to blame. [Barthes has killed off the authors but I would like to know in that case why dictators still periodically sentence them to death]. The Greek word for this – mimesis – used to be translated as "imitation" before the less literal "representation" supplanted it. But I feel that 'imitation' captures something of the suspicion and trepidation with which artists were regarded by their first and harshest critics. Both Plato and Aristotle remarked on art's method of showing things and interrogated it, coming up with conflicting conclusions on the mechanism by which art works on the mind and the desirability or otherwise of its effects. The quandary can be stated as follows: the artist appears to counterfeit or copy life in order to produce art; is the artist therefore a liar and a socially and morally pernicious factor in human life? Epistemology privileges truth in an objective sense while the truth of art is, or seems to be, subjective. How can both definitions of the truth coexist? It is this division which appears to set an insurmountable barrier between the pursuit of truth as it is found in the "hard" sciences and that which the artist engages in, if indeed there is truth in what the artist

engages in.

Since the Greeks did not have this convenient dichotomy to rely on, namely our modern dualism of art and science, they were saddled with the problem of explaining the paradox. Many observers, critics and otherwise, have accused Plato of being a poet, but he denied poets any share in his ideal society, seeing them as subversive and potentially disruptive of the state. He seems to have been less intolerant of the plastic arts, perhaps because the method of sculpture and painting (at least of the Greek kind) is that of straightforward depiction. Like a map or a model, a picture or a statue (or an Urn) does not seek to be anything other than a metonymic emblem of a thing, even a fictional thing like a god. Paintings and sculpture can't *lie*. Actors and poets and texts, however, can. Merely to assert "I am a liar" is to involve oneself in a logical paradox. Words are therefore in themselves tricky, and those whose business it is to mess with words are even trickier.

Aristotle conceded art's anarchic potential, but considered that good art, by arousing and then dissipating socially unacceptable feelings, operated as a safety valve. Using a term, "catharsis", which is also to be found in Greek medical texts and means roughly "purging", he posited a view of art as medicine. Plato too used the word pharmakon to represent the text: a word that can mean either "poison" or "cure". If a modern critic were to stand up today and declare that really good art is the next best thing to a cataclysmic bowel movement, he or she would probably be banned from the world's features pages, or possibly presented with the Turner prize. This is not to say that Aristotle was right, but simply to show how the solidification of disciplinary boundaries over the centuries has allowed the problem to be buried at the crossroads of the trivia. Nearly all modern criticism can be boiled down to a statement on the lines of "All art is true, for a given value of truth". Tacitly assuming that we all know what we are talking about, criticism from the author formerly known as "Longinus" on to modern times has focused on the "how" of art on an *operational* level: how does the author produce this or that effect, how do we see one text as "better" than another. The search for truth has been left to the philosophers, who have generally avoided complicating the already problematic issue of truth and lies in the real world with the bizarre phenomenon of truth and lies in worlds created inside someone's head.

Jacques Derrida, in *Dissemination*, returns to Plato and places him beside Mallarmé in an attempt to reconnect mimesis and poiesis to

truth and falsehood. From what I have understood from reading Derrida, he does not deny in the case of things expressed in language that there is a truth "out there". What he says is that there is a truth "in here", in the text, which moulds what we think of as "true" with regard to the world we live in. The source of the hegemony of the text as sign is that it is powerful enough to be taken for the thing. [Tourists going to see "Juliet's balcony" in Verona; wanting to visit "Hamlet's castle" in HelsingĪr; how many children had Lady Macbeth?] Philosophy itself depends on texts, and philosophers depend on reading. But reading, much more than the physical object of the book, is evanescent, subjective and sometimes in error. Different readings, even by the same person, may produce different truth results. If this is true of works which aim to describe reality, then how much more dangerous are work which do not – namely works of fiction. But Derrida's critique is on the level of Logos; it treats of the phenomenon of language itself, and does not differentiate between the word aspiring to reflect the "real" world and the word aspiring to construct an imaginary one. Most of his instances are taken from non-fiction or from poetry or works like Mallarmé's *Mimique*; he has little to say of *realistic* fiction. Bakhtin, on the contrary, celebrates the novel, but he does not subject it to the kind of scrutiny that Derrida applies to the truth and falsity of the sign. There is therefore a shadow zone between these two levels of analysis, where neither one kind of truth is exclusively operating, and where the phenomenon of literary fiction has its being. However, unlike the shadow zone between quantum and Newtonian mechanics, we have no theorists working on narrowing the gloom; at what point does reality (and lying) cease and fiction begin?

When is a lie not a lie? Scholars may have tacitly agreed that it is bad form to ask this question, but the naïve non-scholar asks it all the time. As Jonathan Culler observed in his 2001 preface to his 1981 classic *The Pursuit of Signs*, about the relative failure of semiotics to arrive at a coherent understanding of signification, "[W]e have not satisfactorily answered the basic questions about how we identify plots, how we recognize satisfactory endings and so on. Critics are more interested in interpreting novels than in trying to spell out how we go about understanding them as we read." Authors use the real world in their writing; they cannot avoid doing so. They transmute it by a process which we call creativity, but the salient features of the reality substrate remain, or are intensified. Things remain the same, only more so. Think for example of Arundhati Roy's *The*

God of Small Things, in which many of the characters are based on real people. Think of George Orwell's *Animal Farm* which gets its kicks ripping off a certain kind of social ideology, or Milan Kundera's novels and their picture of Czech life. Most of us have no difficulty identifying these as fictions and treating them as such. We don't take them seriously, they're *stories* – yet they have in them something that reorganizes and produces a constructed coherence out of the real world: a coherence that is *not* ideologically neutral and is, in some sense that we tie ourselves in knots to explain, more filled with "poetic" truth. Now think of *The Satanic Verses*. This book was banned because the Islamic clergy felt it contained lies about Islam – a particular kind of "truth" which they felt themselves to be guardians of, but it was also criticized by liberal scholars on the grounds of distorting historical fact. The men who banned that book were not literary theorists. They were reacting to the elements of caricature and hyperbole that no one will deny exist in the book. For most of us, namely those who are neither priests nor historians, the fact that the book was a self declared fiction *excused* the presence of those elements, elements that would have been unacceptable in a history of Islam or a treatise on racism in Britain. Most of us felt that the book proscribers were missing the point, but nevertheless we sensed their moral discomfort with what the book said. Is it not significant that a euphemism for lying is "telling tales"? *Animal Farm* and *The Joke* are lies that *might* be true. That is why they frighten people. Pathological liars often create strange and wonderful pasts for themselves which in detail and realization can be in some cases be compared with the works of professional fiction writers. Writers, conversely, often see themselves as "sick", or are sick, but resist cure because they fear the cure will rob them of their function as writers. To be a writer, they secretly believe, one must be a little weird, a little bit of a pathological liar. Anyone who has written fiction will know how scary is the feeling of a process that pulls out of one the stuff of one's living mind, memory, secret loves and hates, and turns it into art. Art has no respect for social conventions; it has a short way with Freudian repression in the mind of the writer. The writer is given a choice: produce art that is valid, vital, vivid, or hang on to the shreds of your respectability or your illusions about yourself and others.

But this can have catastrophic consequences for those who are the "truth source" of the writer's inspiration. To know someone, and then to find that the image of you in their eyes has emerged from

their head as a fictional character, is an even more frightening experience than to create the fictional character. A student of any author's life will find stories of friends who cut the writers dead after they appeared fictionalized in stories. In extreme cases the results have been libel suits.

To take another example out of Rushdie, the character of Raman Fielding in *The Moor's Last Sigh* is transparently based on Bal Thackeray. If Rushdie had published a newspaper article describing the character of "Raman Fielding", he might well have been liable to action for libel. However, had Thackeray sued Rushdie for writing the book, he would have been on very sticky ground.

But if we concede that the only difference between fiction and lying is one of perception and reception, then where does that leave art and its producers and consumers? Would anyone in this room accept a world where all writing was to be assessed on the basis of severely literal truth, where not only the author of fiction, but the reader, the publisher, the critic and the memorable fictional character, were ruled out of court? But if we don't accept such a world, how do we differentiate fiction from the pernicious lie, not just in our own eyes but in the eyes of our critics, of those who do not sympathize with fiction? And if we are to avoid such a bonfire of the vanities, then is it not essential that we describe, in terms that cross disciplinary boundaries, what it is about fictional truth (oxymoron) that is so "true"? The importance of this lies in the fact that literature does not belong to theorists alone. Authors are human beings subject to the same political and economic constraints as the rest of us. Once a fiction is published, no one, not even its publishers, can control what happens to it, who reads it and what they find in it. If it falls into the hands of powerful individuals who wish to see it as a lie and who believe that lying deserves punishment, then its production will be punished, unless there are ways and weapons to defend it.

REFERENCES

Culler, Jonathan. *The Pursuit of Signs.* London: Routledge, [1981] 2001.

Derrida, Jacques. *Dissemination.* Translated in Christopher Norris, *Derrida.* London: Fontana, 1987.

———. *Of Grammatology,* trans. Gayatri Chakravorty Spivak. Boston: Johns Hopkins University Press, 1976.

———. *Writing and Difference,* trans. Alan Bass. London: Routledge, [1978] 2001.

List of Contributors

ARIZPE, PAULA
Institute of Technologico Autonomo de Mexico,
Rio Hondo #1 Tizapan San Angel Mexico,
DF Departmento Academico de Estudios Generales,
Mexico.

ARSTILA, VALTTERI
Research Scholar,
Dept. of Philosophy,
University of Turku,
Turku, Finland.

BORGE, STEFFEN
Dept. of Philosophy
Syracuse University
Syracuse, New York, USA.

BRINCK, INGAR
Dept. of Philosophy & Cognitive Science,
Lund University,
Kungshuset Lundagard,
Lund, Sweden.

CHANDRASEKHARAN, SANJAY
Ph.D. Scholar,
Cognitive Science,
Carleton University,
Ottawa, Ontario, Canada.

CHATTERJEE, RIMI. B.
Asst. Professor,
Dept. of Humanities & Social Sciences,
Indian Institute of Technology,
Kharagpur, India.

Chaudhari, Supriya
Professor,
Dept. of English,
Jadavpur University,
Kolkata, India.

Dass, Tulsi
Professor
Department of Physics,
Indian Institute of Technology
Kanpur, India.

Gardenfors, Peter
Dept. of Philosophy & Cognitive Science,
Lund University,
Kungshuset Lundagard,
Lund, Sweden.

Johnston, William
Senior Lecturer,
Faculty of Humanities and Social Sciences,
Manchester Metropolitan University, UK.

Khan, Abrahim. H.
Graduate Centre for study of Religion
Trinity College,
University of Toronto,
Canada.

Kundan
Cultural Integration Fellowship,
California Institute of Integral Studies,
San Francisco, USA.

Menon, Sangeetha
Associate Fellow, Consciousness Studies,
Philosophy of Science Unit,
National Institute of Advanced Studies,
Indian Institute of Science Campus,
Bangalore, India.

Murphy, Dominic
Asst. Professor,
Division of Humanities & Social Sciences,
California Institute of Technology,
Pasadena, California, USA.

Patnaik, B. N.
Professor,
Dept. of Humanities & Social Sciences,
Indian Institute of Technology,
Kanpur, India.

Perry, James F.
Honors Professor of Philosophy,
Hillsborough Community College,
Tampa, Florida, USA.

Phillips, Stephen
Professor,
Dept. of Philosophy,
University of Texas
Austin, Texas, USA.

Rao, Shobini L.
Additional Professor
Dept. of Clinical Psychology
National Institute of Mental Health and Neurosciences
Bangalore, India.

Ray, G. C.
Professor & Head,
Department of Electrical Engineering,
Indian Institute of Technology
Kanpur, India.

Sanyal, Indrani
Professor,
Department of Philosophy
Jadavpur University
Kolkata, India.

Saunders, B. A. C,
Professor, Department of Social and Cultural Anthropology,
K.U. Leuven, Tiensestraat, Leuven,
Belgium.

Sudarshan, E. C. G.
Professor,
Department of Physics,
University of Texas,
Austin, USA.

TANDON, P. N.
President,
National Brain Research Centre Society,
New Delhi, India.

VAN BRAKEL, J.
Higher Institute of Philosophy
Catholic University of Leuven,
Kardinaal Mercienrplein 2,
Leuven, Belgium.

VATSYAYAN, KAPILA
President,
India International Centre
New Delhi, India.

Walter, Sven.
Research Assistant,
Philosophisches Seminar
Department of Philosophy,
University of Saraland,
Germany.